Chapter 1

General introduction: Tailor-made α-glucans by GH70 glucansucrase enzymes

In preparation for submission

Introduction

Fossil resources are currently the major energy source and primary feedstock for the chemical industry. However, these resources are finite and unsustainable. At the same time, the widespread use of fossil resources causes severe environmental problems, including climate changes and air pollution. The transition from fossil-based economy to bio-based economy has been proposed to be one of the solutions (1-3). Nowadays, the concepts of bio-based economy and bio-refinery have received more and more attention from all over the world (2). In the bio-refinery, large quantities of renewable biomass are used as raw material to produce fuels and chemicals through direct extraction, enzymatic transformation and fermentation (1-5). Products developed from renewable biomaterials possess special functional properties, such as being environmental-friendly, biocompatible, non-toxic and biodegradable (6). Carbohydrates are the largest class of organic compounds and account for about three-quarters of all biological compounds in nature (3,7-9). Carbohydrate compounds are generally synthesized by plants and microbes through photosynthesis and are represented by the formula $C_n(H_2O)_m$.

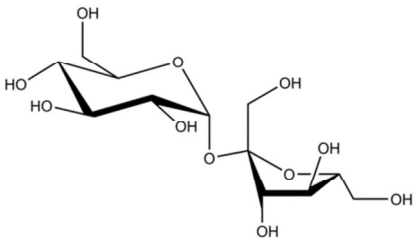

Figure 1. Molecular structure of sucrose (α-D-Glc*p*-(1↔2)-β-D-Fru*f*)

Sucrose, also known as table sugar, is one of the most abundant carbohydrates consumed in our daily life. It is a disaccharide with the formula $C_{12}H_{22}O_{11}$, consisting of the monosaccharides glucose (D-glucopyranose) and fructose (D-fructofuranose), linked by an (α1↔β2) glycosidic linkage (Fig. 1). It is mainly extracted from plants, such as sugar cane and sugar beet. The worldwide production of sucrose is about 170 million tons in 2014 (10). It is universally used as a sweetening agent and is an important component of our daily diet (11). Sucrose from sugar cane is also industrially fermented to produce ethanol as biofuel (11). Alternatively, sucrose may serve as a starting material for the

synthesis of value-added products through fermentation, including industrial chemicals, oligosaccharides and biopolymers for applications such as surfactants, gelling and adhesive agent (6,11-14).

Lactic acid bacteria (LAB) are known to produce extracellular polysaccharides (EPS) from sucrose. Due to the "Generally Regard As Safe" (GRAS) status of LAB, their EPS have found valuable applications in the food industry (15-18). Depending on their composition, EPS are divided into two groups: heteropolysaccharides and homopolysaccharides (19). Heteropolysaccharides of LAB contain different types of monosaccharides (e.g. glucose, galactose and rhamnose), while homopolysaccharides of LAB consist of only one type of monosaccharide (glucose or fructose). The biosynthesis of heteropolysaccharides is complex and requires combined actions of a large number of proteins including enzymes, transporters and regulators (19-21). Generally, the Leloir glycosyltransferase enzymes that are involved require expensive nucleotide-activated sugars (e.g. UDP-glucose) for the synthesis of heteropolysaccharides. On the contrary, homopolysaccharides are generally synthesized from sucrose using a single glucansucrase or fructansucrase enzyme (19,22). These are non-Leloir type of glycosyltransferases, using the cheap and abundantly available sucrose as substrate. Similar to nucleotide-activated sugars, the glycosidic bond of sucrose also has high energy content (~27.6 kJ/mol) (23). Glucansucrase and fructansucrase are efficient enzymes for the synthesis of glucosides or fructosides *in vitro*. Both glucansucrase and fructansucrase catalyze the cleavage of sucrose in the first step. However, glucansucrases use the glucosyl moiety of sucrose for the synthesis of α-glucans, whereas fructansucrases use the fructosyl moiety of sucrose for the synthesis of β-fructans (24). Fructansucrases of LAB have been reviewed elsewhere (25-27). The current review will focus exclusively on glucansucrase enzymes.

Glucansucrases of glycoside hydrolase family 70 (GH70) catalyze the synthesis of α-glucans, which differ in size, type of linkages, and degree of branching. These features influence the physico-chemical properties of the polysaccharides (24,28). The α-glucans produced by glucansucrases from sucrose have (potential) applications in the food, medicine and cosmetic industries. Hence, synthesis of α-glucans could be one of the avenues to produce value-added products from

sucrose. This report surveys the recent developments in glucansucrase research with emphasis on their microbiological distribution, catalytic mechanism, structure-function relationship, reaction specificity and linkage specificity.

Microbiological distribution of glucansucrases

Glucansucrases are mostly found in LAB, such as *Leuconostoc*, *Streptococcus*, *Lactobacillus* and *Weissella* (29). LAB are known to produce lactic acid from sugar metabolism and have been used in the fermentation of food for ages. Various strains also are proven probiotics due to their beneficial effects on human health (30). The ability of LAB to produce various exopolysaccharides has raised new interest in view of its strong potential for industrial applications. Glucansucrases are extracellular enzymes and depending on the particular bacterial source, they are produced either as cell wall-attached or free enzymes in culture fluids, or both (31-33). A diversity of glucansucrases has been characterized from various LAB and were found to produce α-glucans with all the possible glycosidic linkages [(α1→2), (α1→3), (α1→4) and (α1→6)], each enzyme with its own linkage specificity. With the fast development of genome sequencing, the number of glucansucrases annotated is rapidly increasing. By June 2015, in total 266 glucansucrases had been annotated, including 60 characterized in the family GH70 of Carbohydrate-Active Enzymes Database (CAZy). Glucansucrases are mainly found within the genera *Leuconostoc* (63 of 266), *Streptococcus* (153 of 266), and *Lactobacillus* (24 of 266). Some glucansucrases are also present in other LAB, i.e. *Weissella* (21 of 266) and *Oenococcus* (1 of 266). Four genes from non-LAB were also annotated as glucansucrase enzymes, i.e. *Exiguobacterium* (2 of 266), *Bacillus* (1 of 266) and *Azotobacter* (1 of 266), the only Gram-negative source. Some LAB strains produce more than one glucansucrase enyme. For example, *Streptococcus mutans*, which is the main pathogen responsible for dental caries, produces three distinct glucansucrases (GTFB, GTFC and GTFD) (Table 1) (34,35). Six different glucansucrases (DSRA, DSRB, DSRE, DSR-DP, DSR-M and BRS-A) are found in the genome of *Leuconostoc mesenteroides* NRRL B-1299 (Table 1) (36). Several other strains i.e. *Leuconostoc mesenteroides* NRRL B-1355 (Table 1) (37) and *Streptococcus sobrinus* were also found to contain multiple glucansucrases (38). These multiple glucansucrase enzymes generally display different product (linkage) specificity (Table 1) but it has remained unclear whether they have

different physiological roles. Recently, many α-glucan-producing LAB strains have been isolated from fermented food or sugar syrups and were found to possess glucansucrases. In a recent study, a total of thirty LAB from French traditional sourdoughs have been screened for the diversity of exopolysaccharides produced from sucrose (31). These LAB are mainly *Leuconostoc* and *Weissella* strains. They were found to produce glucans with various glycosidic linkage [(α1→2), (α1→3) and (α1→6)] and the presence of glucansucrase-enconding genes was confirmed.

Glucansucrase enzymes from the genus *Leuconostoc*

Early in 1861, Pasteur found a microorganism-derived substance being responsible for the gelification of sugarcane syrups and it was named "dextran" later on (39,40). Van Tiehem isolated this microorganism and named it as *Leuconostoc mesenteroides* in 1878 (39,40). An enzyme from the cell-free supernatant was found to be responsible for the synthesis of dextran (41,42). Now, dextran is defined as a homopolysaccharide which is composed of D-glucose residues with (α1→6) linkages in the main chain and different degrees of (α1→2) or (α1→3) branched linkages. The enzyme that synthesizes dextran is named accordingly as dextransucrase (EC 2.4.1.5).

Leuconostoc is most often found in fermented food. Glucansucrase enzymes are widespread in *Leuconostoc* and the expression of glucansucrase from *Leuconostoc* is generally induced by sucrose (43). Using chemical mutagenesis, mutants (*L. mesenteroides* B-512FMC, B-742CA, B-742CB, B-1142C, B-1299C, B-1355CA and B-1355CB), constitutively expressing glucansucrases, were obtained from wild-type strains *L. mesenteroides* NRRL B-512FM, B-742, B-1142, B-1299 and B-1355 (43-45). These mutant strains produce glucans with identical structures to those of the wild-type organisms. Most of the glucansucrases from *Leuconostoc* strains produce dextran with mainly (α1→6) linkages and minor (α1→3) branch linkages (Table 1). The dextran produced by the glucansucrase DSRS from *L. mesenteroides* NRRL B-512F has been studied most and is widely used in medicine, food and cosmetic industry (46). Other α-glucans with different structures are also produced by *Leuconostoc* bacteria. For instance, DSRE from *L. mesenteroides* NRRL B-1299 is a novel enzyme that synthesizes dextran with a large amount of (α1→2) branched linkages (Table 1)

(47). The molecular characterization of this enzyme showed the presence of two catalytic domains (CD1 and CD2), separated by a central glucan-binding domain (47). Biochemical studies showed that CD1 catalyzed the synthesis of the glucan main chain with predominantly (α1→6) linkages, whereas CD2 formed (α1→2) branched linkages on the (α1→6) main chain (47-49). *L. mesenteroides* NRRL B-1299 also produces two additional glucansucrases (DSRA, DSRB). DSRA synthesizes an α-glucan with both (α1→6) and (α1→3) linkages (Table 1) (50), while the α-glucan produced by DSRB contains larger amounts of (α1→6) linkages (Table1) (51). The genome sequence analysis of *L. mesenteroides* NRRL B-1299 strain revealed the presence of three more glucansucrases (DSR-DP, BRS-A and DSR-M) (36). DSR-DP and DSR-M mainly catalyze the synthesis of (α1→6) linkages while BRS-A catalyzes the synthesis of (α1→2) branched linkages (36). Interestingly, *L. mesenteroides* NRRL B-1355 produces a glucansucrase (ASR) (Table 1) (37,52), synthesizing a glucan with alternating (α1→6) and (α1→3) linkages (52). This distinct α-glucan has been named alternan and its enzyme as alternansucrase (ASR, EC 2.4.1.140). Recently, more α-glucan-producing *Leuconostoc* bacteria and corresponding glucansucrases have been isolated from various sources. An isolate from sugarcane juice was identified as *Leuconostoc citreum* B/110-1-2 encoding a novel dextransucrase (DSRF) (53). DSRF synthesizes a dextran with 93% (α1→6) linkages, 6% (α1→3) linkages and 1% (α1→4) linkages (Table 1) (54). A glucansucrase DSRI of *L. mesenteroides* NRRL B-1118. was recently characterized and produced an insoluble α-glucan with approximately 50% (α1→6) and 50% (α1→3) linkages (55). In a recent report, four and five putative glucansucrase genes were identified in the genome sequences of *L. citreum* LBAE-E16 and LBAE-E16, respectively (56).

Glucansucrase enzymes from the genus *Streptococcus*

Streptococcus strains, especially *S. mutans*, have been recognized as the major dental caries pathogenic bacteria (57,58). Dental caries is generally initiated by biofilm formation involving extracellular polysaccharides produced by microorganisms (58,59). One of the main components of this biofilm is α-glucan (10-20% dry weight of biofilm) (60,61). The biofilms also trap other bacteria and food debris. Once established, the bacteria in the biofilm ferment sugars and produce acids that cause dental caries (57,62). The glucansucrases from

Streptococcus are constitutively expressed (43). *S. mutans* produces three distinct glucansucrases (GTFB, GTFC and GTFD) (Table 1) (34,35,63,64). GTFB (formerly known as GTF-I) and GTFC (GTF-SI) synthesize water-insoluble glucans with large amounts of (α1→3) linkages [designated as mutan and its corresponding enzyme as mutansucrase (EC 2.4.1.125), while GTFD (GTF-S) catalyzes the synthesis of water-soluble glucan with mainly (α1→6) linkages (34,35). It has been reported that the inactivation of any of the three enzymes resulted in a decrease of smooth-surface carious lesions in the specific-pathogen-free rat model system (65). In another study, it was demonstrated that GTFB and GTFC played an important role in cellular adherence to smooth surfaces (66). However, deletion of *gtfD* only slightly affected its cellular adherence. Using GTF-deficient *S. mutans* mutants, it was shown that the presence of all GTFs at optimum ratio was important for sucrose-dependent adherence (35). *S. sobrinus* 6715, which is involved in dental caries as well, also contains multiple glucansucrases, producing soluble and insoluble glucans (38). *Streptococcus downei* Mfe28 also produces two glucansucrases GTF-S and GTF-I (Table 1), being responsible for the synthesis of soluble glucan and non-soluble glucan, respectively (67). Considering the importance of GTFs in the process of dental caries, it has been suggested that specific inhibitors of GTF enzymes may be effective for preventing dental caries (35,58).

Glucansucrase enzymes from the genus *Lactobacillus*

Lactobacillus strains are widely spread in nature and have been used for food application for ages. Some species, e.g. *Lactobacillus reuteri* strains, are considered as probiotic strains due to their beneficial effects for human health (68). *Lactobacillus* strains are found to produce α-glucans by novel glucansucrases. In a previous study, a total of 182 *Lactobacillus* strains were screened for EPS production with sucrose rich medium, 60 of them were found to produce EPS (glucan or fructan), of which 17 produced large amounts (more than 100 mg/L) (69). The glucansucrases from *Lactobacillus* are expressed constitutively (24,33,70). Later on, the genes encoding these glucansucrases were cloned and the enzymes were biochemically characterized (70-72). GTFA from *L. reuteri* 121 synthesizes an α-glucan with 58% (α1→4) linkages and 42% (α1→6) linkages (Table 1) (70). This was the first report about a GH70 glucansucrase producing an α-glucan with (α1→4) linkages from sucrose. α-Glucan containing

large amounts of (α1→4) linkages is referred to as reuteran and its corresponding enzyme as reuteransucrase (EC 2.4.1.-). Structural analysis of reuteran produced by GTFA revealed a large amount of alternating (α1→4) and (α1→6) linkages (73). GTFO of the probiotic strain *L. reuteri* ATCC 55730 represents another reuteran-producing glucansucrase and its reuteran has an even larger amount of (α1→4) linkages (~80%) compared to GTFA (Table 1) (74). A variety of other GTFs producing dextran (GTF180, GTFKg15, GTFKg3 and GTF33) and mutan (GTFML1) are also identified in the genus *Lactobacillus* (Table 1) (71). For example, GTF180 from *L. reuteri* 180 produces an α-glucan with 69% (α1→6) linkages and 31% (α1→3) linkages, while GTFML1 synthesizes an α-glucan with large amounts of (α1→3) linkages (~70%).

Although sucrose remains the canonical D-glucosyl donor substrate for glucansucrases, some enzymes also can use oligosaccharides as donor substrates (75). In this so called disproportionation reaction, the glucosyl group is transferred from one saccharide donor to an identical or similar saccharide acceptor substrate (75). Binder *et al*. found that dextransucrase from *L. mesenteroides* NRRL B-512F and GTF-S from *S. mutans* 6715 has disproportionation activity with isomalto-oligosaccharides, malto-oligosaccharides and panose. The closely related enzyme amylosucrase from family GH13 has also been shown to have disproportionation activity with malto-oligosaccharides (76). GH13 represents the largest family in CAZy database, with mainly α-amylase enzymes that are active on (α1→4) glucan substrates. GH70, together with GH13 and GH77 (4-α-glucanotransferase), form clan GH-H. CGTase (family GH13) (77) and α-glucanotransferase (GH77) (78) are examples of well-characterized disproportionating enzymes in clan GH-H.

This disproportionation activity of glucansucrases has been overlooked for decades due to its relatively low activity compared to the glucosyl transfer from sucrose. Recently, a new subfamily of glucansucrases (GTFB from *L. reuteri* 121, GTFW from *L. reuteri* DSM 20016 and GTFML4 from *L. reuteri* ML1) (Fig. 2) were identified as 4,6-α-glucanotransferases, which were inactive with sucrose and showed disproportionation activity with (α1→4)-containing malto-oligosaccharides and starch (79,80). The 4,6-α-glucanotransferase enymes from these *L. reuteri* strains showed mainly (α1→4) to (α1→6) transferase activity

and produced linear isomalto/malto-oligosaccharides mixtures with an increasing percentage of (α1→6) linkages (80,81). These 4,6-α-glucanotransferase enzymes show about 50% amino acid sequence identity with other glucansucrase enzymes and clearly belong to family GH70. In view of their structural similarity with family GH70 and their reaction similarity with family GH13 using malto-oligosaccharides as substrates, 4,6-α-glucanotransferases are considered to be a novel evolutionary intermediate of families GH13 and GH70 (79,80). It is worth to note that in the respective genomes both the GTFB and GTFML4 encoding genes are located directly upstream of the regular glucansucrase enzymes GTFA and GTFML1, respectively, while only GTFW (not another regular glucansucrase) was identified in *L. reuteri* DSM 20016 (71,72,79,80). It remains unknown whether this adjacent location has a physiological role.

Glucansucrases from other lactic acid bacteria

Apart from *Leuconostoc, Streptococcus* and *Lactobacillus*, other LAB (i.e. *Weissella*) also contain genes encoding glucansucrase enzymes. *Weissella*, originally included in the genus *Leuconostoc*, also produces EPS from sucrose. Most of the α-glucans produced by *Weissella* strains are dextrans with large amounts of (α1→6) linkages (more than 90%) (Table 1). *Weissella confusa* 39-2 and *Weissella cibaria* LBAE K39 were isolated from wheat sourdoughs and found to produce linear dextrans (82). The glucansucrase (DRSC39-2) from *W. confusa* 39-2 has been characterized (83). *W. cibaria* isolated from human saliva was found to encode a glucansucrase (DSRWC), synthesizing a soluble glucan with large amounts of (α1→6) linkages (84). Also a putative glucansucrase gene from *Oenococcus oeni* PSU-1 and two from *Exiguobacterium* (non-LAB) were identified by genome sequencing (www.cazy.org), but these genes have not been characterized.

Table 1. Examples of glucansucrase enzymes characterized from different Lactic Acid Bacteria and the glycosidic linkage composition of their α-glucan products

Species	Strains	Enzymes	Glucans	Linkage composition (%)				References
				(α1→6)	(α1→3)	(α1→4)	(α1→2)	
Leuconostoc	*mesenteroides* NRRL B-512F	DSRS	dextran	95	5			(85)
	mesenteroides NRRL B-1299	DSRA	dextran	85	15			(50)
		DSRB	dextran	95	5			(51)
		DSRE	dextran	81	10	3	5	(49)
		DSR-DP	dextran	100				(36)
		DSR-M	dextran	100				(36)
		BRS-A*	(α1→2)				34	(36)
	mesenteroides NRRL B-1355	ASR	alternan	57	43			(52)
	citreum B/110-1-2	DSRF	dextran	93	6	1		(54)
	mesenteroides NRRL B-1118	DSRI	mutan	50	50			(55)
Streptococcus	*mutans* GS5	GTFB	mutan	12	88			(63)
		GTFC	mutan	15	85			(64)
		GTFD	dextran	70	30			(34)
	oralis	GTFR	dextran	86	14			(86)
	downei Mfe 28	GTF-S	dextran	90	10			(67)
		GTF-I	mutan	12	88			(87)
Lactobacillus	*reuteri* 121	GTFA	reuteran	42		58		(72)
	reuteri ATCC 55730	GTFO	reuteran	21		79		(74)
	reuteri 180	GTF180	dextran	69	31			(71)
	reuteri MLI	GTFMLI	mutan	35	65			(71)
	sakei Kg15	GTFKg15	dextran	90	10			(71)
	fermentum Kg3	GTFKg3	dextran	92	8			(71)
	parabuchneri 33	GTF33	dextran	81	19			(71)
Weissella	*confusa* 39-2	DSRC39-2	dextran	97	3			(83)
	cibaria	DSRWC	dextran	100				(84)

*With sucrose as sole substrate, BRS-A does not catalyze polymer synthesis. In the presence of linear dextran (33 mM) as acceptor substrate and 146 mM sucrose as donor substrate, 34% (α1→2) linkages were found in the product mixture (36).

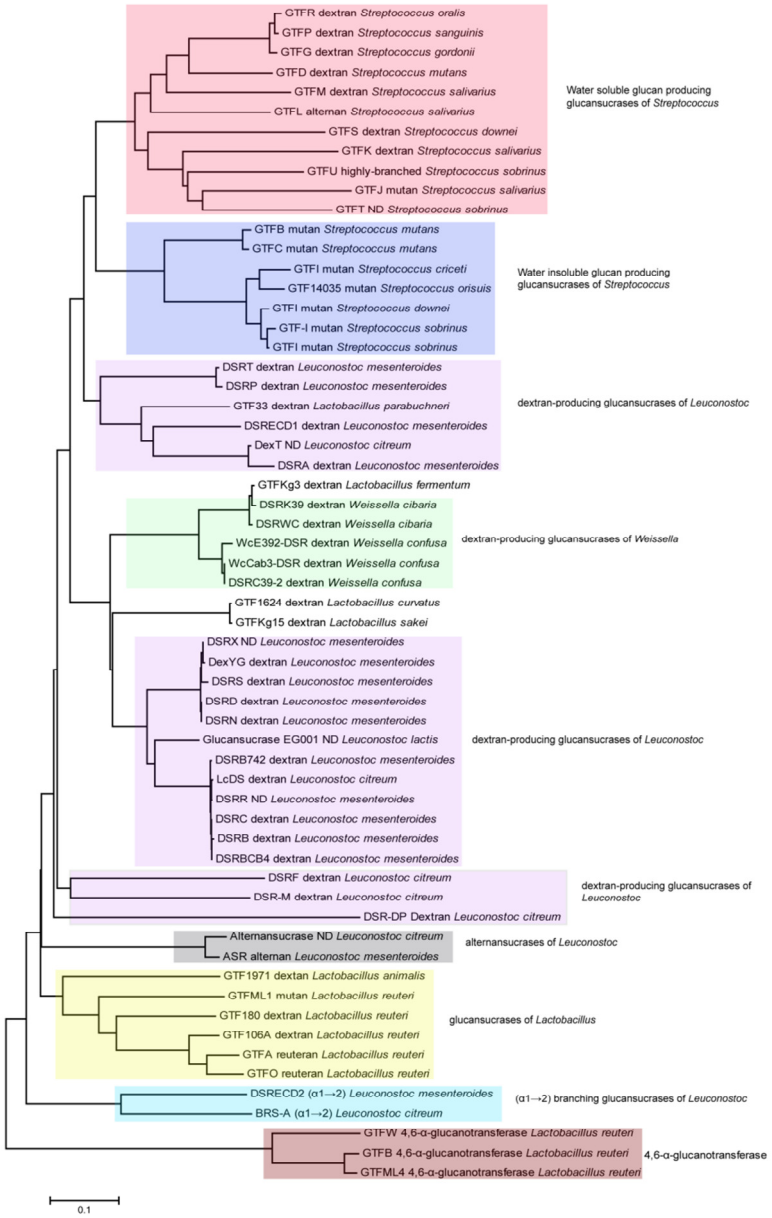

Figure 2. Phylogenetic tree of all characterized GH70 glucansucrase enzymes. Each sequence is labelled with the enzyme, α-glucan polysaccharide synthesized and the name of the bacterial species. Details of aligned sequences were shown in Table S1. MEGA 6.0 was used to perform alignment and phylogenetic analysis (using neighbor joining method). Only the catalytic cores of glucansucrase enzymes (domains A, B and C, see

below) were used for the analysis. ND: not determined. The bar represent a genetic distance of 0.1 substituion per position.

Phylogenetic analysis

Phylogenetic analysis of the biochemically characterized glucansucrase enzymes using their catalytic core sequences only (domain A, B and C, see below) shows that most of these cluster in clades reflecting their bacterial hosts, but with a few exceptions (Fig. 2). Most glucansucrase enzymes of *Leuconostoc* produce dextran type [mainly (α1→6)] of polysaccharides and they form three separate clusters through the tree. The *Leuconostoc* alternansucrase and (α1→2) branched glucansucrase form separate clades from the main cluster of *Leuconostoc* glucansucrase enzymes, reflecting their separate product specificity. Glucansucrase enzymes from *Streptococcus* form two main clades, which produce soluble (dextran) and insoluble (mutan) polysaccharides, respectively. Glucansucrase enzymes of *Weissella* are closely related and produce dextran types of polysaccharides. Also *Lactobacillus* derived glucansucrase enzymes cluster together, however, they produce different types of polysaccharides (reuteran, dextran and mutan). It is worth to note that the three *Lactobacillus* 4,6-α-glucanotransferase enzymes clearly differ from the other LAB glucansucrase enzymes, in line with their proposed evolutionary intermediate status between families GH13 and GH70 (79,80).

Structures of glucansucrase enzymes

Primary structure

Glucansucrases are large proteins with an average molecular weight of ~160 kDa. The primary structure of virtually all glucansucrase proteins shows the same organization with only a few exceptions. The amino acid sequences of different glucansucrases contain four different regions (Fig. 3): i) signal peptide (SP), ii) N-terminal variable region (VR), iii) conserved catalytic domain (CD), and iv) C-terminal glucan-binding domain (GBD) (71,88,89). All glucansucrases contain a typical signal peptide (36 to 40 amino acids) of gram-positive bacteria in their N-termini, reflecting their extracellular location.

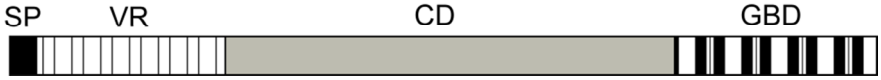

Figure 3. General primary structure of glucansucrase proteins from Lactic Acid Bacteria. SP: signal peptide; VR: variable domain; CD: catalytic domain; GBD: glucan binding domain.

Following the signal peptide, a variable region (200 to 700 amino acids) is present. The variable region of glucansucrases from *Leuconostoc* and *Streptococcus* is relatively short (~ 200 amino acids), while glucansucrases (GTFA, GTF180, GTFO and GTFML1) from *L. reuteri* strains contain large variable regions (~700 amino acids) (71,72,90). Different repeat units have been identified in the N-terminal variable regions (Table 2). For example, glucansucrase DSRS and DSRB, and alternansucrase ASR from *L. mesenteroides* contain A-repeats (Table 2) in their N-termini (90). Motif T and S have been identified in DSRT from *L. mesenteroides* NRRL B-512 F (91) and in DSRE from *L. mesenteroides* NRRL B-1299 (47), respectively. Glucansucrases (GTFA, GTF180, GTFO and GTFML1) from *L. reuteri* possess five RDV repeats (Table 2) in their relatively large N-termini (70,71). The other glucansucrases from *Lactobacillus* contain different repeat units. GTF33 from *L. parabuchneri* 33 contain a unique repeat unit-TTQ (Table 2) while GTFKg3 from *L. fermentum* and GTFKg15 from *L. sakei* have several less-conserved YG repeats (Table 2) (71). The functions of these N-terminal variable regions are still unknown. DSRA from *L. mesenteroides* NRRL B-1299 has no N-terminal variable region (50), suggesting that this region is not essential for enzyme activity. Indeed, deletion of N-terminal variable region has no significant effect on the products formed and activities of glucansucrases (92). Deletion of the N-terminal variable region of GTFA even significantly increased its initial transferase activity of the purified enzyme, but had no effect on the structure of α-glucan synthesized (70).

The GBD also have been shown to contain different amino acid sequence repeat units, which were divided into four classes based on the sequence similarity: A, B, C and D repeat (Table 2). These repeats were also present in the glucan-binding protein of *S. mutans*, toxin A of *Clostridium difficile* and the lysins of *Streptococcus pneumonia* (24,28). Within these repeats, a consensus YG repeat

was identified which is characterized by the presence of a cluster of aromatic amino acids and a glycine residue three or four residues downstream of the aromatic cluster (93). Compared with the GBD of glucansucrases from *Leuconostoc* and *Streptococcus*, the GBD of glucansucrases from *L. reuteri* (GTFA, GTFO, GTF180 and GTFML1) are relatively short. Moreover, they lack A, B, C and D repeats. Instead, several conserved and less-conserved YG repeats are present in their GBD (51,71,88,94). Sequential deletion from the C-terminus onwards yielded mutants that displayed decreased glucan binding ability and decreased activity (70,88). For instance, GTFA mutants with C-terminal sequential deletion have impaired glucan binding ability (70). The hydrolytic activity of these mutants did not significantly change, while their transferase activity was reduced. Therefore, GBD was proposed to be involved in glucan binding and possibly in polysaccharide chain growth.

Table 2. Repeat units in the N-terminal variable region (VD) and the C-terminal glucan binding domain (GBD) of glucansucrase enzymes

Repeat units	Sequences	References
A-repeats	WYYFNXDGQAATGLQTIDGQTVFDDNGXQVG	(87)
B-repeats	VNGKTYYFGSDGTAQTQANPKGQTFKDGSVLRFYNLEGQYVSGSGWY	(87)
C-repeats	GKIFFDPDSGEVVKNRFV	(67)
D-repeats	GGVKNADGTYSKY	(95)
Motif T	TDDKA(A/T)TTA(A/D)TS	(91)
Motif S	PA(A/T)DKAVDTTP(A/T)T	(47)
RDV	R(P/N)DV-$_{X12}$-SGF-$_{X19-22}$-R(Y/F)S	(72)
TTQ	TTTQN(A/T)(P/A)NN(S/G)N(D/G)PQS	(71)
YG	NDGYYFXXXGXXH^{O}X(G/N) H^{O}H^{O}H^{O}	(93)

X indicates non-conserved amino acids; H^{O} represents hydrophobic amino acids.

Based on sequence similarity, glucansucrase enzymes are classified as members of glycoside hydrolase family 70 (GH70) in the CAZy database. Secondary structure prediction demonstrated that the conserved catalytic domain of GH70 enzymes contains the $(\beta/\alpha)_8$ barrel as found in the closely related enzymes from

family GH13 (α-amylase, cyclodextrin glucanotransferase and amylosucrase) and family GH77 (4-α-glucanotransferase) (24,96,97). The $(\beta/\alpha)_8$ barrel is characterized by the presence of 8 β-strands (β1-β8) residing in the core of the enzyme, alternating with 8 α-helices (α1-α8) surrounding the β-strands (Fig. 4). Members of GH clan H (families GH13, GH70 and GH77) share the catalytic $(\beta/\alpha)_8$ barrel. However, unlike family GH13 and GH77 enzymes, the $(\beta/\alpha)_8$ barrel of glucansucrases is circularly permutated. From N-terminus to C-terminus, the circularly permuted $(\beta/\alpha)_8$ barrel of GH70 glucansucrases starts with the α-helix that corresponds with α3 of family GH13 enzymes with the sequence of N-terminus-α3-β4-α4-β5-α5-β6-α6-β7-α7-β8-α8-β1-α1-β2-α2-β3-C-terminus (Fig. 4A and 4B) (97).

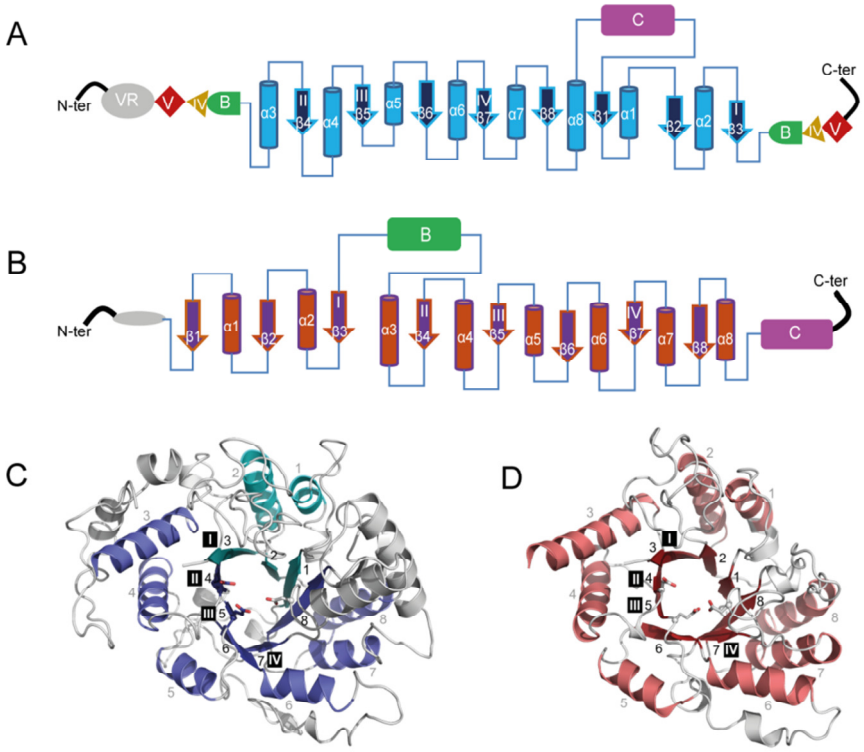

Figure 4. Topology diagrams models of family GH70 glucansucrases circularly permutated $(\beta/\alpha)_8$ barrel (A) and the family GH13 α-amylase (B) $(\beta/\alpha)_8$ barrel. Cylinders represent α-helices and arrows represent β-strands. The equivalent α-helices and β-sheets in GH70 glucansucrases and GH13 α-amylases are numbered the same. The different domains in GH70 and GH13 enyzmes are indicated. Domain C of GH70 glucansucrase is inserted between α-helix 8 and β-strand 1 while that of family GH13 α-amylase locates at

the C-terminus. Domain B of GH13 α-amylases is inserted between β-strand 3 and α-helix 3 while that of GH70 glucansucrases is formed by two discontinuous polypeptide segments from both the N- and C-terminus. The same is true for domains IV and V of GH70 glucansucrases. A variable region (VR) is present in the N-terminus of GH70 glucansucrases. The four conserved regions (I-IV) which are located in β-strands 3, 4, 5, and 7, respectively and are shared between family GH70 and GH13 enzymes, are shown in the figure. Folds of the $(\beta/\alpha)_8$ barrels in the GH70 representative GTF180-ΔN (C, PDB: 3KLK) of *L. reuteri* 180 and in the GH13 representative α-amylase of *Bacillus licheniformis* (D, PDB: 1BPL) are presented. Only domain A constituting the $(\beta/\alpha)_8$ barrel is shown and $(\beta/\alpha)_8$ barrel and colored for a better representation. α-Helice and β-strands are numbered, and the conserved regions (I-IV) are indicated at the corresponding β-strand. The circularly permutated $(\beta/\alpha)_8$ barrel of GH70 glucansucrases is formed by two separate polypeptide segments (N-terminal parts colored in deepblue and C-terminal parts colored in cyan), which is caused by the insertion of domain C.

Prior to the availability of crystal structures of glucansucrases, their catalytic mechanism was mainly explored by comparative studies with GH13 enzymes. The four conserved regions (I to IV) of the GH13 family enzymes, with the 3 catalytic residues and various residues interacting with bound donor and acceptor substrates, are also present in glucansucrases (70,98). Conserved regions I-IV locate at β strands 3, 4, 5 and 7, respectively. Due to the circularly permutated structure of glucansucrases, their region I is located C-terminal of regions II to IV (Fig. 4 and 5). Six of the seven conserved residues from region I to IV in family GH13 are also present in family GH70 (24). Only the His122 (Taka-amylase A numbering) in family GH13 is replaced by Gln (Gln1509, GTF180 numbering) in family GH70, which was shown to be involved in transition state stabilization in family GH70 (99). This suggests that the mechanism of glycosidic linkage cleavage by GH70 glucansucrase enzymes is similar to that of GH13 family α-amylase enzymes. The catalytic triad of GH70 glucansucrase has been identified by sequence alignment with GH13 enzymes and confirmed by mutagenesis studies. Asp1025 (GTF180 numbering) acts as nucleophile and is involved in formation of the covalent glucosyl-enzyme intermediate (100). Mutagenesis of the corresponding residue in different glucansucrases completely suppressed enzyme activity (70,85,100). Glu1063 is the acid/base catalyst which donates a proton to facilitate the leaving of fructose and abstracts the proton from the hydroxyl group of the acceptor substrate to activate it for attacking the glucosyl-enzyme intermediate (24,70,100). The third catalytic residue (Asp1136) acts as transition state stabilizer (100).

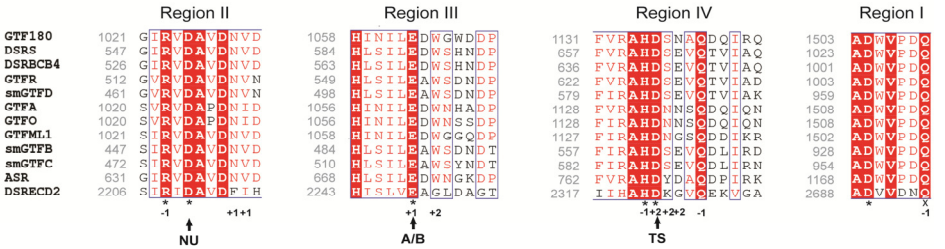

Figure 5. Sequence alignment of homology regions I-IV of family GH70 enzymes. NU: nucleophile; A/B: acid/base catalyst; TS: transition state stabilizer. -1, +1 and +2 indicate residues interacting with the donor substrate sucrose and the acceptor substrate maltose in the crystal structure of GTF180-ΔN with sucrose or maltose (see below) (100). Symbol * designates six of seven conserved residues from regions I to IV in family GH70 that are also present in family GH13. Symbol x indicates the seventh conserved residue (Q) from region I in family GH70 enzymes, which is replaced by a H residue in family GH13 enzymes. GTF180: GTF180 (Uniprot; Q5SBN3) from *L. reuteri* 180; DSRS: DSRS (Q9ZAR4) from *L. mesenteroides* NRRL B-512F; DSRBCB4: DSRBCB4 (D2CFL0) from *L. mesenteroides* B-1299CB4; GTFR: GTFR (Q9LCH3) from *S. oralis* ATCC10557; smGTFD: GTFD (P49331) from *S. mutans* GS5; GTFA: GTFA (Q5SBL9) from *L. reuteri* 121; GTFO: GTFO (Q4JLC7) from *L. reuteri* ATCC 55730; GTFML1: GTFML1 (Q5SBN0) from *L. reuteri* ML1; smGTFB: GTFB (P08987) from *S. mutans* GS5; smGTFC: GTFC (P13470) from *S. mutans* GS5; ASR: ASR (Q9RE05) from *L. mesenteroides* NRRL B-1355; and DSRECD2: DSRE catalytic domain 2 (Q8G9Q2) from *L. mesenteroides* NRRL B-1299. The aligned sequences were prepared with EsPript (101) and modified according to (29).

3D structures of glucansucrase enzymes

Although large efforts have been made to generate crystals of glucansucrase proteins suitable for X-ray diffractions studies, only recently the first 3D structure of glucansucrase proteins have been solved. To date, the crystal structures of four GH70 glucansucrases [GTF180-ΔN from *L. reuteri* 180, GTFA-ΔN from *L. reuteri* 121, GTF-SI (amino acid residues 244-1163) from *S. mutans* and ΔN_{123}-GBD-CD2 of DSR-E from *L. mesenteroides*] have been determined, revealing a common domain organization (48,100,102,103). In all cases, a truncated enzyme (i.e. lacking the N-terminal variable region, but remaining full activity) was used for crystallization. The crystal structures of these truncated glucansucrases have provided us with important information about domain organization, catalytic mechanism and specificity. Rather surprisingly, the 3D structures of these glucansucrases revealed a novel domain organization (Fig. 5), different from the one predicted on the basis of sequence alignments (Fig. 3). The polypeptide

chains of the truncated glucansucrases follow a U-shape path to form five domains (A, B, C, IV and V) (Fig. 6) (100). Except for domain C, each of the four domains is formed by two discontinuous polypeptide chains from both the N- and C-termini. The catalytic core consists of three domains (A, B and C), which resembles that of family GH13 enzymes.

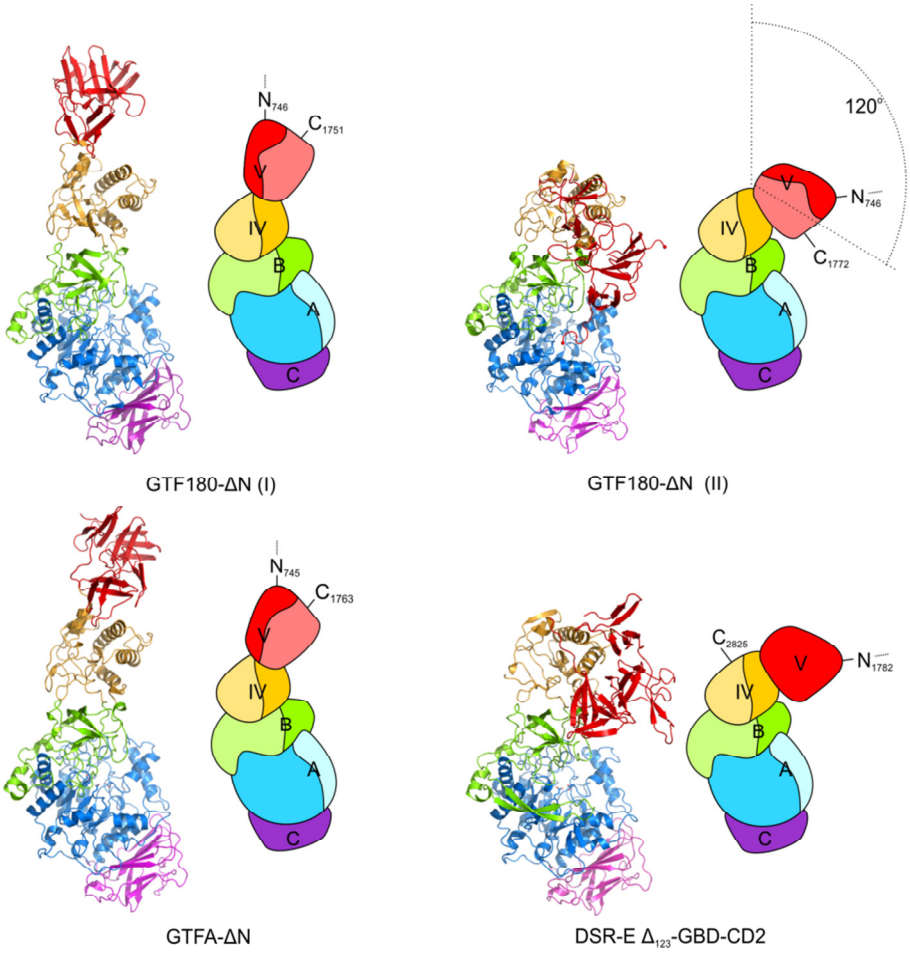

Figure 6. Three-dimensional structures and schematic domain organization of glucansucrases from family GH70. Different domains are colored in blue (A), green (B), magenta (C), yellow (IV) and red (V). Crystal structures [GTF180-ΔN I (PDB: 3KLK, 1.65 Å), GTF180-ΔN II (PDB: 4AYG, 2.0 Å), GTFA-ΔN (PDB: 4AMC, 3.60 Å) and DSR-E ΔN_{123}-GBD-CD2 (PDB: 3TTQ, 1.90 Å)] are shown. This figure has been adapted from (29).

Domain A contains the circularly permutated $(\beta/\alpha)_8$ barrel as predicted in contrast to the $(\beta/\alpha)_8$ barrel in GH13 enzymes (Fig. 4C and 4D) (97). The four conserved regions (I to IV), including the seven conserved residues, also reside in domain A (Fig. 4C) (24,100). Specifically, the three catalytic residues (the nucleophile Asp1025, acid/base catalyst Glu1063 and transition state stabilizer Asp1136, GTF180 numbering) are located at loops following β-strands β4, β5 and β7, respectively (Fig. 4C) (100). Two discontinuous inserts with a large stretch of amino acids between β-strand 3 and α-helix 3 from both N- and C-termini form a separate domain B next to domain A (Fig. 4 and 6). The position of domain B in GH70 glucansucrase enzymes is similar to that in GH13 enzymes. The active site of glucansucrases is located in a pocket-shaped cavity and lies at the interface of domain A and domain B (100). Domain B contributes several amino acids (L938, L940, A978 and L981) for shaping the substrate/acceptor binding sites, which may determine the enzyme product specificity. (100). Domain C of glucansucrase enzymes, located at the bottom of the U-shape (Fig. 6), contains an eight-stranded β-sheet with a Greek key motif, which is similar to that of domain C in family GH13 enzymes (100,104). It is the only domain that is formed by a continuous polypeptide segment. The function of domain C remains unclear, although it is widely distributed within the GH13 and GH70 families.

In addition to domains A, B and C, glucansucrases from family GH70 have two extra domains (IV and V) attached to the catalytic core. Domain IV lies between domain B and domain V. The structure of domain IV is novel and it has no similarity to any other known protein structure and only occurs in glucansucrase proteins (29,100). Domain V is located next to domain IV. It contains several sequence repeats (Table 2) which have been shown to be involved in glucan binding (24,90,93,94,105). Structural analysis of domain V revealed the presence of a consensus β-solenoid fold with multiple copies (29,100). The precise roles of domains IV and V have remained unknown. It has been proposed that domain IV acts as a hinge that facilitates the glucan chain growing by bringing the glucan chain bounded with domain V toward and away from the catalytic site, but no experimental evidence is available yet (102). Comparison of crystal structures of different glucansucrases revealed a positional variability of domain V (Fig. 6). For example, compared to the crystal structure of GTF180-ΔN (I), domain V of GTFA-ΔN showed a shift around 20 Å with respect to the other domains (103).

Surprisingly, domain V of ΔN_{123}-GBD-CD2 of DSRE is located in a completely different position and swings down to compact with the catalytic core (48). Although domain V of ΔN_{123}-GBD-CD2 of DSRE takes a different position, it adopts a similar fold as domain V of GTF180-ΔN (I) (48). A B-factor analysis of domain V in the different crystal structures showed that the average B-factor of domain V is higher, which indicates that it is more flexible (106). Recently, the flexibility of domain V was also demonstrated by the elucidation of a new crystal form of GTF180-ΔN (II) with a 120^{o} rotation at a hinge located between domains IV and V (Fig. 6), further supported by the observation of positional flexibility of domain V in solution (106). Truncation of domain V from GTF180-ΔN did not have significant effects on enzyme activity (107). However, higher amounts of oligosaccharides were produced at the cost of the polysaccharide production (107). This provided direct evidence for the involvement of domain V in polysaccharide synthesis.

Although the crystal structure of a full length glucansucrase is not available yet, small angle X-ray scattering studies have showed that the N-terminal variable region (~ 700 amino acids) extends further away from domain V (106). As a result, the overall shape of GTF180 showed an almost symmetric boomerang-like molecular shape with the bend point located between domain IV and V.

Catalytic mechanism of glucansucrase enzymes

Similar to family GH13 enzymes, the α-retaining double displacement reaction mechanism is used by glucansucrases (24,88,104). This two-step mechanism involves 3 catalytic residues, a nucleophile, a acid/base catalyst and a transition state stabilizer (Fig. 7). In the first step, the (α1↔β2) glycosidic linkage of sucrose is cleaved by the attack of the nucleophile with the formation of a β-glucosyl-enzyme covalent intermediate. This glucosyl-enzyme intermediate is stabilized by the transition state stabilizing residue. The acid/base catalyst protonates the fructosyl moiety, resulting in release of fructose. In the subsequent step, the glucosyl moiety is transferred to the non-reducing end of an acceptor with retention of the α-anomeric configuration. The next reaction cycle starts again to synthesize gluco-oligosaccharides and α-glucan polysaccharides from sucrose. The crystal structure of GTF180-ΔN validates that glucansucrases use the same set of amino acids to catalyze the reaction as family GH13 enzymes

(100,104). The crystal structure of the inactive mutant GTF180-ΔN D1025N bound with sucrose revealed that of the seven strictly conserved (catalytic) residues, six also employed by GH13 enzymes, make similar interactions with the -1 glucosyl moiety of sucrose (Figs. 5 and Fig. 8). The glucosyl moiety binds at the -1 site by the conserved interaction with R1023, D1025, H1135, D1136, E1063, Y1465 and Q1509, of which R1023, D1025, H1135, D1136 and Q1509 make direct H-bonds to glucosyl hydroxyl groups. Residue D1025 and E1063 are oriented towards the glycosidic oxygen and the anomeric C1 atom of the glucosyl moiety of sucrose. Residue D1025 acts as the nucleophilic residue which attacks the anomeric C1 carbon of the glucosyl unit of sucrose to form a β-glucosyl-enzyme covalent intermediate, stabilized by the transition state stabilizer (D1136) and residue E1063 is the acid/base catalyst donating a proton to facilitate the release of fructose and depronating the acceptor molecule to activate it. Residue Y1465 locates at the bottom of subsite -1, providing hydrophobic interactions with the glucosyl moiety of sucrose. The conserved residue D1504 is outside of the catalytic center, making a H-bond to the conserved Y1465 hydroxyl group. Furthermore, the subsite -1 is lined by residues Q1140 and N1411, which form hydrogen bonds with the C3 hydroxyl group of the glucosyl moiety through a water molecule. This results in a pocket-shaped active site which can only accommodate one glucosyl moiety, demonstrating that glucansucrases are able to transfer only one glucosyl unit in each reaction cycle. In contrast, a cleft-shaped active site with multiple donor subsite is present in family GH13 α-amylases. In GTF180-ΔN, the fructosyl moiety binds more loosely at the +1 site, having interactions with residues E1063, W1065, N1029, D1136 and Q1140 (Fig. 8). The C3 hydroxyl group of the fructosyl moiety is involved in a hydrogen bond network with the side chains of W1065, D1136 and E1063 while N1029 and Q1140 make a hydrogen bond with the C1 and C6 hydroxyl groups, respectively. In addition, two hydrophobic residues (L981 and L982) from domain B make van der Waals interactions with the fructosyl moiety.

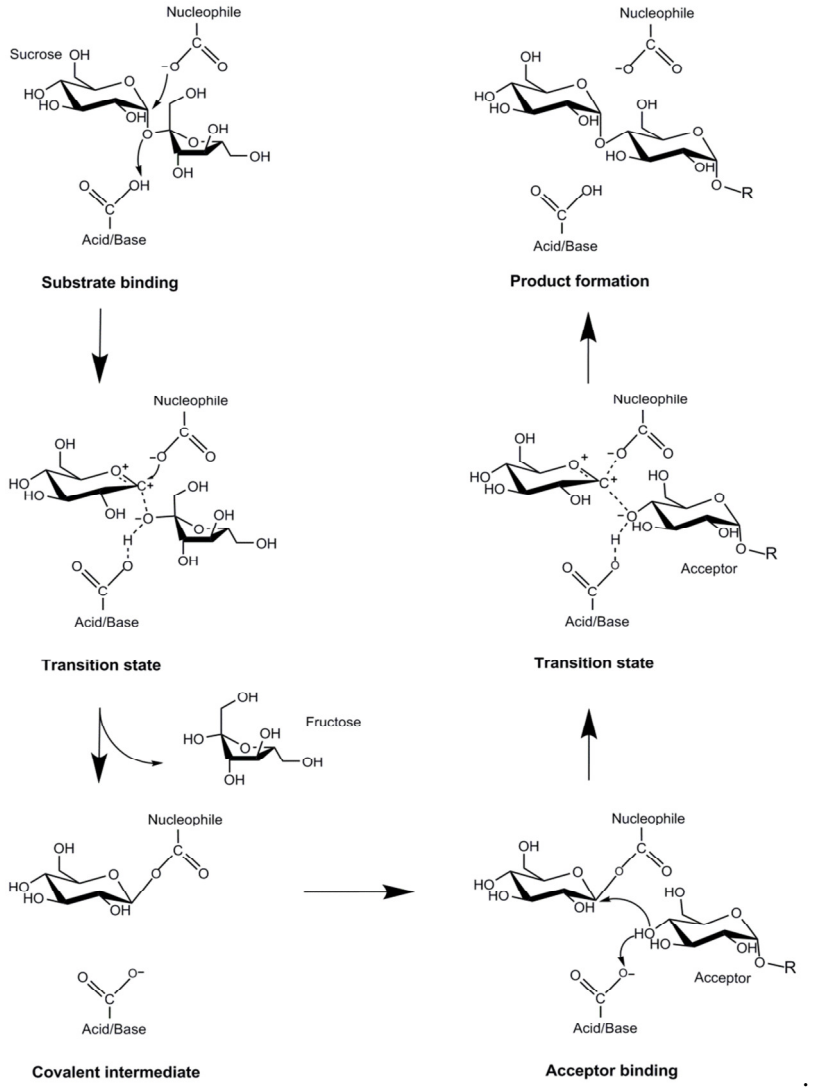

Figure 7. Non-reducing end elongation mechanism for the synthesis of α-glucan by glucansucrase. This figure has been adapted from (29). In the glucansucrase enzymes, the nucleophile is an Asp (D1025 in GTF180), the acid/base catalyst is a Glu (E1063) and a transition state stabilizer (Asp, D1136) is also involved. In the catalysis cycle shown in this figure, an (α1→4) glycosidice linkages is formed.

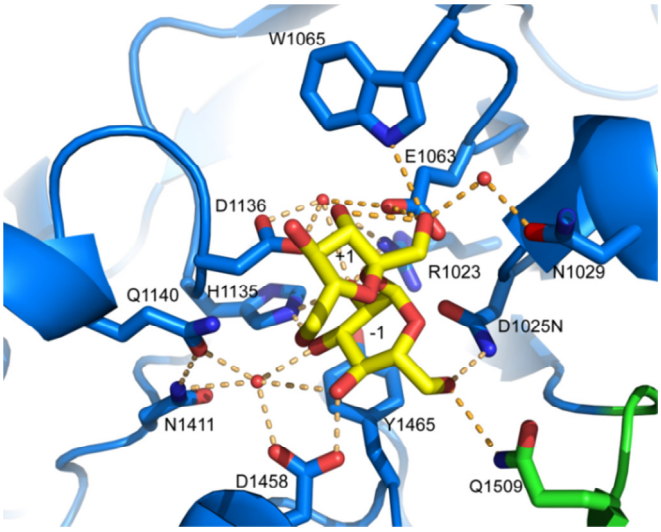

Figure 8. Sucrose (donor substrate) binding site -1 and +1 in the crystal structure of the GTF180-ΔN mutant D1025N sucrose complex (3HZ3) (100). Sucrose is shown with yellow carbon atoms. Residues from domain A (blue) and B (green) surrounding the -1 and +1 subsites are indicated. Hydrogen bonds are shown as dashed lines.

Reactions catalyzed by glucansucrases

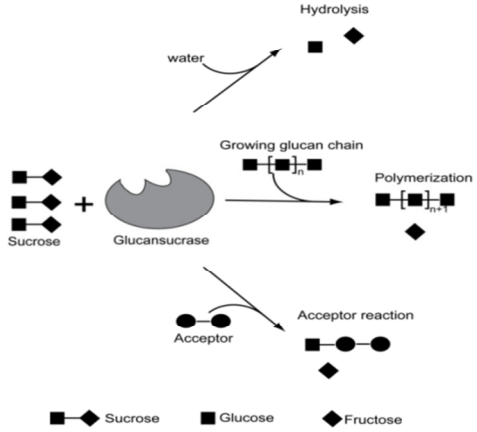

Figure 9. Reactions catalyzed by glucansucrase enzymes using sucrose as donor substrate and various acceptor substrates.

The glucosyl-enzyme covalent intermediate is formed in the first step of the catalytic cycle. Depending on the acceptor substrate used in the second step,

glucansucrases catalyze three different reactions, namely polymerization, hydrolysis and acceptor reactions (Fig. 9) (24,28,29). In the polymerization reaction, a growing glucan chain is used as acceptor substrate leading to the elongation of the glucan chain with one more glucosyl unit. Water can also be used as acceptor substrate, resulting in hydrolysis of sucrose into glucose and fructose. In the acceptor reaction, the glucosyl group from sucrose is transferred to a carbohydrate or non-carbohydrate acceptor substrate containing a hydroxyl group to produce an oligosaccharide or glucoside.

α-Glucan synthesis

In spite of the availability of crystal structures, the mechanism of glucan synthesis by glucansucrase is not fully understood yet, especially regarding the initiation of α-glucan synthesis, the synthesis process, the formation of branches, the mode of action and the linkage specificity.

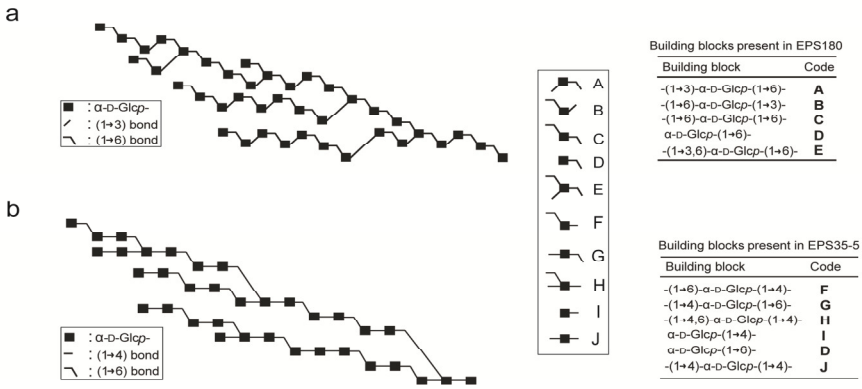

Figure 10. Visual representation of the composite models of the α-glucan polysaccharides produced by GTF180 (a) (108) and by GTFA (b) (73).

An intriguing question is how α-glucan synthesis by glucansucrases is initiated, starting from sucrose. In other words, how does the α-glucan chain grow during the synthesis process. The structures of the GTFA and GTF180 α-glucan polysaccharides have been examined by 1D and 2D NMR spectroscopy, together with methylation analysis. Additionally, structural analysis has been performed on isolated oligosaccharides obtained by enzymatic hydrolysis, partial acid hydrolysis and Smith degradation of polysaccharides. (73,108-110). Composite models of the α-glucans produced by wild-type and mutant glucansucrases were

constructed by combining all the information derived from the above mentioned analysis (Fig. 10). These models provide valuable information about the structures of these α-glucans and the structure-function relationships of GTFA and GTF180 glucansucrase enzymes. Since glucansucrases are able to transfer only one glucose unit in each reaction cycle, it is expected that α-glucans are synthesized by step-wise addition of one glucosyl unit to the growing glucan chains (24,100). Characterization of the oligosaccharides formed in the early phase of the reaction in time therefore would provide important information about the synthesis process. For that purpose, oligosaccharides formed by GTFA during the initial phase of the reaction, were isolated and structurally characterized (111). The results showed that the main oligosaccharides were sucrose elongated with alternating (α1→4) and (α1→6) linkages (111). The abundance of these linkages is also shown in the composite model of the α-glucan synthesized by GTFA (Fig. 10). This demonstrated that polysaccharide synthesis starts with the transfer of glucosyl units to the non-reducing glucose end of sucrose with the formation of alternating (α1→4) and (α1→6) linkages. Subsequently, once the oligosaccharides reach a certain degree of polymerization (DP), branched linkages are formed and are further elongated with alternating (α1→4) and (α1→6) linkages. As a result, a branched α-glucan with large amounts of alternating (α1→4) and (α1→6) linkages is synthesized. This study provides direct evidence that sucrose is used as a primer by GTFA in the synthesis of α-glucan. The low molecular mass (DP 20-30) and linear dextran produced by GTF-S3 from *S. sobrinus* was also found to be terminated with a sucrose moiety (112). Remaud-Simeon *et al.* also reported that DSRS of *L. mesenteroides* NRRL B-512F and alternansucrase of *L. mesenteroides* NRRL B-1355 used sucrose as initiator of polysaccharide synthesis (88). However, a series of isomalto-oligosaccharides (acceptor products starting from glucose) is also formed reaching DP higher than 25 by DSRS of *L. mesenteroides* NRRL B-512F, while oligosaccharides with sucrose at the reducing end did not exceed a DP higher than 12 (88). In this work, both sucrose and glucose were proposed as initial acceptor for polysaccharide synthesis; but the latter was preferred (88). However, whether this could be extended to other glucansucrases is not known yet.

Most of the α-glucans synthesized by glucansucrases are branched, to varying degrees (24,28,29). Most glucansucrases do not require an extra enzyme for the

formation of branches in their α-glucan products. The mechanism for forming branched linkages remains unclear. Robyt and Taniguchi proposed that the formation of branched linkages is through the acceptor reaction of glucansucrases (113). Site-directed mutagenesis has identified several amino acid residues that are involved in the formation of branched linkages (114). Mutation of three residues (S1137:N1138:A1139) following the transition state stabilizer residue (D1136) in GTF180 resulted in several mutants which synthesized α-glucans with a higher degree of branches (109). These residues are located close to the +2 glucosyl unit of maltose in the crystal structure of GTF180-ΔN in complex with maltose (Fig. 11) (100). Residues D1085, R1088 and N1089 from α-helix 4 are located at the other side of the +2 glucosyl unit of maltose and they all make an indirect hydrogen bond with the +2 glucosyl unit of maltose through the same water molecule (Fig. 11). Multiple and single mutations in these residues resulted in mutants producing hyperbranched α-glucans (15%-22% branching) (Chapter 7). Single mutation studies showed that D1085 and R1088, but not N1089, are responsible for the increase in branched linkages. Irague *et al*. also reported that mutations in the corresponding residues of DSRS of *L. mesenteroides* NRRL B-512F (D460, H463 and T464) increased the proportion of (α1→3) linkages (115). Mutational studies of A978 and D1028 in GTF180 identified the involvement of a groove preceding the +1 site in the formation of branched linkages (Chapter 6). Partially blocking the groove by mutating these residues to amino acid residues with larger side chains reduced the amount of branched linkages in the α-glucans synthesized. To summarize, the proportion of branched linkages in the α-glucans produced by glucansucrases could be manipulated by mutations close to the acceptor substrate binding subsites. Glucansucrase enzymes indeed may form branched linkages through the acceptor reaction, involving dissociation of the growing glucan chain from the acceptor binding subsites and subsequent rebinding in a different way that allows formation of branched linkages. However, the question, when and where branched linkages are formed, remains unanswered.

There also has been a controversy about the mode of glucan chain elongation (processive or non-processive). Previously, glucansucrases were found to synthesize high molecular weight polysaccharides during the early phase of reaction without the detection of intermediate oligosaccharides (116). Consequently, glucansucrases were assumed to act processively in the synthesis

of α-glucan polysaccharides. Using a more sensitive method (HPAEC-PAD), oligosaccharides were detected later on, which implied that glucansucrases act non-processively (88). However, kinetic analysis of polysaccharide synthesis by glucansucrase DSRS of *L. mesenteroides* NRRL B-512F revealed that high-molecular-mass (HMM) dextrans reached the maximum size after only 23% of sucrose consumption with the simultaneous detection of oligosaccharides (88). The polysaccharide synthesized by GTFA was also found to reach maximum size in a short time with the detection of oligosaccharides (111). The polysaccharide size did not increase further, not even with the availability of excess sucrose (Chapter 3). The detection of HMM polysaccharide with maximum size indicates a processive mode while the detection of oligosaccharides points at non-processive mode. Therefore, taking into account all the information, Remaud-Simeon *et al.* proposed a semi-processive mechanism of polymerization for GH70 glucansucrase enzymes (88). The structural basis for processive polysaccharide synthesis and non-processive oligosaccharide synthesis is proposed to be located in the repeat units in the C- and N-termini, which have been shown to be involved in glucan binding (70,90,117). Mutants of DSRS with truncated A repeats were less efficient in polysaccharide synthesis compared with the wild-type enyme (88). This suggests that these repeats facilitate polysaccharide synthesis by anchoring the growing polysaccharide chain close to the active site. With the availability of the crystal structures, it became clear that the sequence repeats present in domain V formed a modular β-solenoid fold and domain V was proposed to play a role in carbohydrate binding (100,102). Indeed, truncation of domain V from GTF180-ΔN heavily impaired polysaccharide synthesis and increased the oligosaccharide synthesis (107). Interestingly, mutation of residues around the acceptor binding sites partially restored the polysaccharide synthesis of GTF180-ΔNΔV (107). Furthermore, the polysaccharide synthesis increased significantly by mutation of residues around the acceptor binding site (L940) of GTF180-ΔN (118). These results suggest that the structural basis for processive polysaccharide synthesis lies in both domain V, and the acceptor binding sites, representing remote and close binding sites for glucan chains, respectively. The glucan chain may also bind at the protein surface located in between these sites. The elucidation of the crystal structures of glucansucrase with high-molecular-weight glucan chains could reveal more information about these sites. A good example for this is a study of

polysaccharide synthesis by amylosucrase, which is a special glucansucrase from the family GH13. Structural analysis showed that amylosucrase does not possess domain IV and V (119,120). The crystal structure of amylosucrase with maltoheptaose revealed three different oligosaccharide binding sites (OB1, OB2 and OB3) (120). OB1 locates at the active site, OB2 is in close proximity to the B' domain (unique for amylosucrase) which is close to the active site and contains a high content of aromatic residues, and OB3 is part of domain C. These OB sites, especially OB2 at the B' domain of amylosucrase were considered to be pivotal for transferase activity and polysaccharide synthesis (120,121). Mutation studies, targeting residues at OB1, confirmed its importance for polysaccharide synthesis and OB2 was shown to provide an anchoring platform for the polysaccharide (119,122). Further molecular modeling studies proposed a model for the mechanism of polysaccharide synthesis involving OB1 and OB2 (122).

The intriguing mechanism of linkage specificity raised more attention and will be discussed in the following sections.

Hydrolysis reaction

Following the initial reaction phase of DSRS of *L. mesenteroides* NRRL B-512F with sucrose, using high sensitivity detection method [High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection (HPAEC-PAD)], it was demonstrated that catalysis starts with hydrolysis of sucrose resulting in detection of glucose and fructose (88). With the availability of sucrose and initially produced oligosaccharides as acceptor substrates, glucansucrases transfer glucosyl moiety of sucrose to acceptor substrates. Glucansucrase then acts as a transferase instead of hydrolase (24,28,29). For glucansucrase, the major determinants for this reaction specificity are still unknown. The structural factors that determine the preference for the transglycosylation reaction have been extensively studied in a related enzyme, Cyclodextrin Glucanotransferase (CGTase) (123). CGTase is a unique member of family GH13 catalyzing the synthesis of circular ($\alpha 1 \rightarrow 4$)-linked oligosaccharides (cyclodextrins) from starch. It has been observed that the protein backbone of CGTase undergoes a small but essential conformational change upon binding of a carbohydrate substrate at acceptor substrate binding site, and donor substrate binding site, enabling the proposal of induced-fit mechanism for CGTase

transglycosylation activity (104,124,125). In this mechanism, binding of sugar substrate at acceptor binding site or at donor binding site would drive the CGTase from one conformation to another conformation, which activates the transglycosylation activity of the enzyme. Structural analysis and mutational studies of CGTase enzymes showed that the hydrophobic amino acid residues in the acceptor binding subsites provide stacking interactions for the carbohydrate acceptor substrate and are important for the transglycosylation activity (126). The hydrosis activity of CGTase can be enhanced by mutating these residues to more polar amino acid residues (126). This mechanism has also been confirmed to be important for transglycosylation activity in several mutagenesis studies targeting other residues at these sites (126-130). The induce-fit mechanism probably is true for glucansucrase enzymes as well. Indeed, a hydrophobic residue (W1065, GTF180 numbering) is located at the acceptor binding site +2 (100). The hydrolysis reaction increases significantly by mutation of this W1065 residue to non-aromatic residues (X. Meng *et al*., in preparation). Moreover, these mutants are also deficient in polysaccharide synthesis. Another residue (N1029) has a direct hydrogen bond with maltose (acceptor substrate) as revealed by the crystal structure of GTF180-ΔN-maltose complex (100). Mutations of this residue also increased the hydrolysis reaction of glucansucrase (Chapter 6).

Acceptor substrate reaction

The acceptor substrate reaction was first reported by Koepsell *et al*. In their study with dextransucrase from *L. mesenteroides* NRRL B-512F, they demonstrated that a large number of sugar and sugar derivatives i.e. maltose, isomaltose, glucose and methyl glucoside can act as acceptor substrates, leading to the production of oligosaccharides at the expense of polysaccharide synthesis (131). Fructose was also found to be used as acceptor to produce sucrose isomers (i.e. leucrose and trehalulose) (24,70,88,131). This generally starts in the later phase of the reaction, probably due to the accumulation of fructose in the reaction digest. Under conditions with a high sucrose concentration, the sucrose isomers were found to be elongated further with glucosyl units (132). Robyt and Eklund investigated the quantitative effects of 17 different saccharide acceptors with dextransucrase from *L. mesenteroides* NRRL B-512F (133). Maltose was found to be the most effective acceptor giving a series of oligosaccharides (DP 3-6), followed by isomaltose, nigerose, methyl α-D-glucopyranoside, 1,5-anhydro-D-

glucitol, D-glucose, turanose, methyl β-D-glucopyranoside, cellobiose and L-sorbose. Lactose, raffinose, melibiose, D-galactose and D-xylose were also found to act as acceptor but with each a single glucosylated product while D-fructose and D-mannose gave two mono-D-glucosylated (disaccharide) products.(133). Nowadays, the acceptor reaction of glucansucrase with carbohydrate acceptors is effectively used to produce novel oligosaccharides from cheap agro-resource sucrose. The other advantage of enzymatic synthesis of oligosaccharides using glucansucrase lies in the tight control of regio- and stereo-specificity, which is tedious with chemical synthesis (114). With maltose as acceptor, alternansucrase produced oligo-alternan series with alternating (α1→6) and (α1→3) linkages (134). The prebiotic effects of different DP components within these oligosaccharides were tested and the results showed that the prebiotic effects were inversely proportional to the size of the oligosaccharides with DP3 possessing the highest prebiotic effects (135). Gentiobiose [β-D-Glc*p*-(1→6)-D-Glc*p*], considered as a potential prebiotic, has a bitter aftertaste which prevents its adoption as food ingredient (136). Gentiobiose was found to act as acceptor substrate of alternansucrase to produce several novel oligosaccharides with reduced or eliminated bitter taste of gentiobiose (137). Lactulosucrose [β-D-Gal*p*-(1→4)-β-D-Fru*f*-(2↔1)-α-D-Glc*p*], a novel oligosaccharide with prebiotic properties, has been shown to be effectively synthesized by dextransucrase from *L. mesenteroides* NRRL B-512F using lactulose [β-D-Gal*p*-(1→4)-D-Fru*f*] as acceptor (138). Isomalto-oligosaccharides (IMO) with controlled size could be obtained from sucrose using dextransucrase of *L. mesenteroides* NRRL B-512F with the addition of maltose or glucose as acceptors (139,140).

Besides carbohydrates, non-carbohydrates, such as aromatic compounds (i.e. catechol), have also been found to act as acceptor substrates for glucansucrases. This characteristic of glucansucrases has been used extensively in chemo-enzymatic synthesis for the glycosylation of a variety of different chemical compounds (114). The glycosylation of a chemical compound holds the potential to improve its physicochemical and biological properties, including the solubility of hydrophobic compounds, the activity of antibiotics, and the flavors of food ingredients (141,142). An enzyme microarray screening system was applied with GTFR from *Streptococcus oralis* to evaluate its acceptor substrate specificity; the results showed that GTFR was able to glycosylate a broad range of alcohols and

amino acid derivatives (143). GTFR has also been used for the efficient preparation of novel branched thiooligosaccharides in a two-step chemo-enzymatic synthesis strategy (144). Different chemical compounds have been tested as acceptor substrates using GTFD from *S. mutans* GS-5. Dihydroxy aromatic compounds like catechol, 4-methylcatechol, and 3-methoxycatechol were found to be efficient acceptor substrates with high yields of glucosylation products (145). The other compounds that have been reported to be glycosylated include flavonoid (146), arbutin (147), salicin (148), and epigallocatechin gallate (149).

Engineering reaction specificity of glucansucrase enzymes

The relative balance of the three reactions (reaction specificity) catalyzed by glucansucrases is different depending on the particular enzyme. It has been shown that the reaction specificity of glucansucrases can be changed by both enzyme engineering and reaction condition engineering. In several studies, truncation of C-terminal sequence repeats shifted the enzyme from polysaccharide synthesis to oligosaccharides synthesis (70,88). It has been shown that truncation of domain V in GTF180 of *L. reuteri* 180 resulted in impaired polysaccharide synthesis and a corresponding increase of oligosaccharide synthesis (107). Several mutants of glucansucrases, targeting amino acid residues close to the acceptor binding site, have been reported to produce different ratios of polysaccharide, oligosaccharide and glucose from sucrose (86,88,118,150,151). For instance, S628 (a residue C-terminal to the transition state stabilizer) mutants of GTFR (S628D and S628R) from *S. oralis* abolished polysaccharide synthesis and only produced short chain oligosaccharides (86). Moulis *et al.* also reported that mutations of the residues C-terminal to the transition state stabilizer of DSRS from *L. mesenteroides* NRRL B-512F and alternansucrase from *L. mesenteroides* NRRL B-1355 abolished or reduced polysaccharide synthesis (88). These residues locate at acceptor substrate binding site +2 and mutations at this subsite may impair the affinity of the enzyme with the growing glucan chain, resulting in impaired polysaccharide synthesis. The hydrolysis could also be altered as discussed above by mutation of residues at acceptor binding site (i.e. W1065, N1029 and L981) (Chapter 6). Interestingly, the L940W mutant of GTF180 showed 6 fold reduced hydrolysis reaction (percentage of sucrose used for hydrolysis reaction) (118). Mutation of leucine to tryptophan probably makes the

active site pocket of GTF180 more hydrophobic and thus reduces the solvent accessibility to the active site (118). Hydrolysis is not preferred in view of the enzymatic synthesis of saccharide. Therefore, reducing hydrolysis is definitely an important aspect of engineering of glucansucrase enzymes; in this light, the L940W mutation provides a promising starting point for further engineering.

Besides enzyme engineering, the reaction conditions, especially sucrose concentration, have been shown to alter the reaction specificity of glucansucrases. Kim, *et al* showed that the amount of HMM dextran produced by dextransucrase from *L. mesenteroides* B-512FMCM decreased while the amount of LMM dextran increased with increasing sucrose concentrations (152). Dextrans of different molecular mass were synthesized at controlled sucrose concentrations, enzyme concentrations and reaction temperatures by dextransucrase from *L. mesenteroides* B-512FMC (153). Lee *et al*. also showed that the chain length of the isomalto-oligosaccharide product produced by DSRS of *L. mesenteroides* B-512F was changed at different ratios of donor substrate (sucrose) and acceptor substrates (maltose) (139). Similarly, the ratio of oligosaccharide synthesis versus polysaccharide synthesis of GTFA from *L. reuteri* 121 is directly proportional to the concentration of sucrose, while the linkage distributions and the sizes of the polysaccharides produced at different sucrose concentrations were identical (132).

Taken together, the reaction specificity of glucansucrases can be engineered by both enzyme and reaction engineering to meet different purposes like the synthesis of oligosaccharides or polysaccharides.

Tailor-made α-glucans by engineering glucansucrase product specificity

Due to their ability to produce a diverse range of α-glucans with different types of linkage, size, branching and hence physico-chemical properties, glucansucrases have attracted interest for industrial applications in food, medicine, cosmetics etc. (46). Dextran produced by the glucansucrase DSRS from *L. mesenteroides* NRRL B-512F is extensively applied as gelling, viscosifying and emulsifying agent in the food industry. Bakery products with dextran have improved softness and increased volume (154). Dextran is also applied as size-exclusion chromatography material in research, and as a plasma expander in

medicine (40,155,156). Various α-glucans have been shown to possess anti-corrosion activity possibly by forming biofilms on the surface of steel (69,157,158). Moreover, α-glucans and oligosaccharides formed by glucansucrases have potential prebiotic activities and therefore can be used to stimulate growth of beneficial intestinal bacteria such as *Bifidobacterium* and *Lactobacillus* (159). Linear and (α1→2)-branched dextran produced by DSRE from *L. mesenteroides* NRRL B-1299 were reported to increase *Bifidobacterium* populations *in vitro* (160). However, the prebiotic activity was not enhanced by the presence of (α1→2) linkages. The effects of glycosidic linkages and molecular masses on prebiotic activity also have been evaluated. Analysis of oligosaccharides with different DP produced by alternansucrase and dextransucrase acceptor reactions showed that the prebiotic effects were inversely proportional to the size of the oligosaccharides, with DP3 possessing the highest prebiotic effects (135,161). Monsan *et al.* reported that the addition of α-glucan to animal feed improved the weight gain of pig and broilers (162). Therefore, the large variety of α-glucans and oligosaccharides with different structures holds great potential for industrial applications. A more detailed understanding of the linkage specificity of glucansucrases may allow production of tailor-made α-glucans with desired properties.

Although the active sites of glucansucrases are highly conserved, they produce α-glucans with different structures especially regarding the glycosidic linkages (24,28,114). Depending on the major linkage in the α-glucans, they are divided into five different groups: (i) dextran, which has a large amount of (α1→6) linkages and is produced by dextransucrases (EC 2.4.1.5, i.e. DSRS) (85,163); (ii) mutan, which contains mainly (α1→3) linkages and is synthesized by mutansucrases (EC 2.4.1.125, i.e. GTF-SI) (92,102); (iii) reuteran, predominantly composed of (α1→4) linkages and produced by reuteransucrase (EC 2.4.1.-, i.e. GTFA) (69,70); (iv) alternan, which is a unique α-glucan containing alternating (α1→6) and (α1→3) linkages and is produced by alternansucrase (EC 2.4.1.140, ASR) (37); (v) the distinct α-glucan containing a large amount of (α1→2)-branched linkages produced (α1→2)-branched glucansucrase (EC 2.4.1.-, i.e. DSRE of *L. mesenteroides* NRRL B-1299) (47). These various α-glucans possess different physico-chemical properties such as molecular mass, solubility and

viscosity. For instance, the solubility of mutan containing mainly (α1→3) linkages is generally low, while dextran with predominantly (α1→6) linkages is more soluble. Dextrans with 3 to 20% of (α1→3) linkages have been produced by DSRS mutants and were shown to have different properties regarding molecular mass and rheological behavior (164).

The linkage specificity of glucansucrase appears to be determined by only a small number of amino acids in view of the highly conserved catalytic residues, (70,114,151). As discussed above, all glucansucrases initiate the reaction cycle by formation of a covalent C1-linked glucosyl-enzyme intermediate (29). Consequently, it has been proposed that it is the way in which acceptor substrates bind at acceptor binding subsites that determines the linkage specificity of a glucansucrase enzyme, revealing the importance of amino acid residues at these sites (70,88,114,151). Prior to the availability of crystal structures, putative regions involved in acceptor binding site have been identified by alignment with family GH13 enzymes in view of their high sequence similarity (88,150,151). These include residues C-terminal to the catalytic nucleophile D1025 (GTF180 numbering, conserved region II, Fig. 5), residues C-terminal to the acid base catalyst E1063 (region III, Fig. 5) and residues C-terminal to the transition state stabilizer D1136 (region IV, Fig. 5). These regions display amino acid variations in different glucansucrases (29). Indeed, mutagenesis studies targeting these residues confirmed their importance for the specificity of glucansucrases. GTFA mutants P1026V and I1029V (region II, GTFA numbering) displayed different product spectra with sucrose as substrate, yielding higher levels of isomaltose and leucrose (151). Mutants A1066N and H1065S:A1066S of GTFA (region III, GTFA numbering) produced similar products as wild-type but displayed lower activity (151). In several glucansucrases, residues located C-terminal to the catalytic transition state stabilizer (N1134:N1135:S1136: Q1137:D1138, region IV, GTFA numbering) have been found to be critical for linkage specificity (86,88,109,151). Combined mutations in the tripeptide (N1134:N1135:S1136, region IV, GTFA numbering) following the transition state stabilizing residue (D1133) in GTFA from *L. reuteri* 121 shifted GTFA linkage specificity from mainly (α1→4) linkages to (α1→6) linkages, indicating its involvement in linkage specificity determination and hence acceptor substrate binding (151). Further mutation studies showed that, among these three residues, residue N1134

plays a major role in linkage specificity determination (150). Similarly, mutations in the corresponding tripeptide in GTFR from *S. sobris* (86), GTF180 from *L. reuteri* 180 (109), and DSRS from *L. mesenteroides* NRRL B-512F (88), altered their linkage specificity as well. Mutation of the residues near the transition state stabilizer in GTFR (R628G:V630I:D717A) resulted in an increase of (α1→3) linkages in the polysaccharide produced (86). Surprisingly, combining the mutations in regions II and IV in GTF180, mutant V1027P:S1137N:A1139S, introduced 12% (α1→4) linkages (not present in the wild-type) in the α-glucans produced (110). Mutants of the fourth residue located at C-terminal transition state stabilizer of GTF180 (Q1140A and Q1140H) produced α-glucans with a higher percentage of (α1→6) linkages. The GTF180 Q1140E mutant also produced α-glucan with 3% (α1→4) linkages (109). In addition, the fifth residue located at C-terminal transition state stabilizer was also shown to be involved in linkage specificity determination. Random mutagenesis of D569 in GTF-I of *S. downei* showed that mutations at this position affected the structure of the α-glucan and the size of the synthesized oligosaccharides (165). Mutations of the fifth residue following the transition state stabilizer in DSRI of *L. mesenteroides* NRRL B-1118 also affected the linkage composition of their products with mostly increased (α1→3) linkages (166). These results demonstrate that the involvement of residues located C-terminal to the transition state stabilizing residue in acceptor substrate binding is a general feature of GH70 glucansucrases. Other residues, further away from the catalytic residues, also have been found to influence the specificity of glucansucrases. For example, mutations T350K and S455K in DSRS of *L. mesenteroides* NRRL B-512F increased the amount of (α1→6) linkages in the α-glucans produced (167). Moreover, the double mutant T350K/S455K was able to produce (α1→2) branches on dextran similar to DSRE.

Elucidation of the crystal structure of glucansucrase GTF180-ΔN with a bound acceptor maltose confirmed that the amino acid residues at the acceptor binding site are critical for linkage specificity and provided structural explanations for the effects of mutation observed previously (48,100,102). Four maltose binding sites (M1, M2, M3 and M4) were revealed by soaking GTF180-ΔN crystals with maltose (100). However, the residues forming the binding site for M2, M3 and M4 are not conserved within family GH70 and only have nonspecific interactions with GTF180-ΔN, indicating that these binding sites are not important. M1 binds

at the acceptor binding site +1 and +2 with its C6 hydroxyl group of non-reducing end glucosyl moiety pointing towards the sucrose binding pocket. An activated C6 hydroxyl group by acid/base catalyst (E1063) would attack the C1 of the glucosyl-enzyme intermediate, resulting in the formation of an (α1→6) linkage. The binding mode of M1 explains how an (α1→6) linkage is formed with maltose as acceptor substrate (Fig. 10). At subsite +1, N1029 makes direct and indirect hydrogen bonds with the +1 C4 and C3 hydroxyl group, respectively; D1028 forms an indirect hydrogen bond with the +1 C4 hydroxyl group; residues from domain B (L938, L940, A978 and L981) shape the groove near the +1 site. These residues have rarely been targeted for mutagenesis studies prior to the availability of crystal structures (114). In the recent site-directed mutagenesis studies, these residues, especially residues from domain B, were found to be critical for linkage specificity and display different roles (118) (Chapter 6 and 7). All the L940 mutants synthesized α-glucans with larger amounts of (α1→6) linkages (118). Surprisingly, the L940W mutant produced linear α-glucan with only (α1→6) linkages; the synthesis of (α1→3) linkages was abolished completely. Docking studies with isomaltotriose showed that this tryptophan blocks a groove, preventing the reducing end of isomaltotriose to occupy the space observed in the wild-type (118). Consequently, the C3 hydroxyl group of the non-reducing end glucose unit is too far away to attack the C1 of the glucosyl-enzyme intermediate. Instead, the C6 hydroxyl group is within the distance for forming (α1→6) linkages. This highlights the critical importance of the groove, where L940 locates, for (α1→3) linkages synthesis in GTF180. Residue A978 was found to be involved in branched linkage formation. Mutations of A978 to residues with a larger side chain (Leu, Pro, Phe and Tyr) reduced the branched linkage formation in the α-glucans produced, while mutations to smaller residues (Gly and Ser) had no significant effects (Chapter 6). All D1028 mutants had increased amounts of (α1→6) linkages in their α-glucan products. The branched linkages were also influenced by D1028 mutations. Again, mutations to residues with a large side chain (Tyr and Trp) reduced the number of branched linkages (Chapter 6). Docking studies showed that both A978 and D1028 are involved in shaping the groove above the +1 site, the space of which is required for acceptor substrate binding for formation of branched linkages (Fig. 12). Mutating these two residues to bulky residues apparently partially blocked the groove, resulting

in a decrease in branched linkage formation. L938 mutants were also shown to produce α-glucans with altered structures and thus also involved in linkage specificity. Mutations of N1029 caused an increase of (α1→3) linkages in the α-glucan products, indicating its involvement in the linkage specificity determination (Chapter 6). At acceptor binding site +2, residues S1137, N1138, A1139, Q1140, D1141, which have been shown to be important for linkage specificity, are located at one side of the glucosyl unit (Fig. 11). It is worth to note that only S1137 has a direct hydrogen bond with the +2 C1 hydroxyl group. Indeed, S1137 was found to be the main determinant for linkage specificity in previous mutation studies (86,88,109,150,151). This confirms the predicted involvement of these residues at this acceptor binding site. At the other side of the +2 glucosyl unit, residues D1085, R1088 and N1089 all possess an indirect hydrogen bond with the +2 C2 hydroxyl group through the same water molecule (Fig. 11). Combined mutations and individual mutations studies showed that these residues are also involved in linkage specificity (Chapter 7). Mutations at these three residues, especially D1085 and R1088, introduced extra (α1→4) linkages in the α-glucans produced. Remarkably, some mutants also produce hyper-branched α-glucans with the proportion of branched linkage from 15% to 22%. Single mutational studies showed that residues D1085 and R1088, but not N1089, were responsible for the increase in branched linkages. In another combined mutagenesis study, targeting the corresponding residues in DSRS also revealed their importance (115). Some residues, which were found to be important for linkage specificity and are located far away from the catalytic residues in the primary structure, were confirmed to participate in shaping the acceptor binding site in the crystal structure, providing a structural explanation for the results observed previously (167). By now, only acceptor binding sites +1 and +2 have been mapped out and no crystal structure of a glucansucrase protein with a bound higher molecular mass acceptor substrate is available; it remains unclear whether amino acid residues at further acceptor binding sites are determinants for linkage specificity. These remote acceptor binding sites are difficult to be identified by sequence alignment or prediction based on crystal structure, because they may be not conserved.

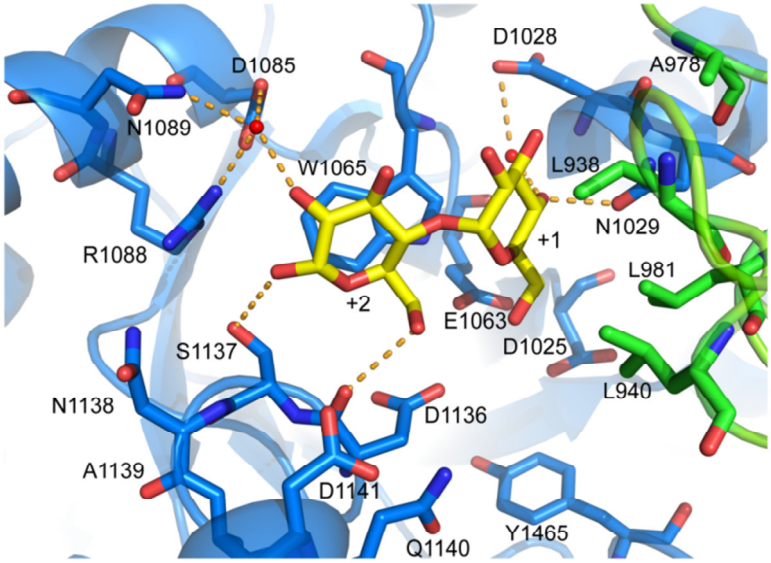

Figure 11. Maltose binding sites +1 and +2 in the crystal structure of the GTF180-ΔN maltose complex (3KLL) (100). Maltose is shown with yellow carbon atoms. Residues from domain A (blue) and B (green) surrounding the +1 and +2 subsites are indicated. Hydrogen bonds are shown as dashed lines.

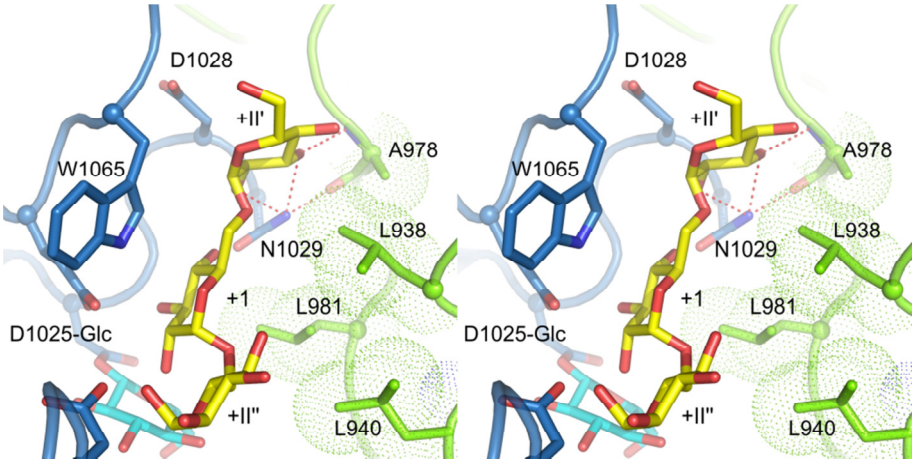

Figure 12. Stereo view of a docked isomaltotriose in the active site of a modeled GTF180-ΔN glucosyl-enzyme intermediate (100). Residues from domain A (blue) and B (green) surrounding the +1 and +II' subsites are indicated.

In some cases, it has been shown that the upcoming linkage type is also determined by the last linkage formed. One example is the alternansucrase (ASR)

from *L. mesenteroides* NRRL B-1355, which catalyzes the synthesis of alternan with alternating (α1→6) and (α1→3) linkages. This suggests that an acceptor substrate with an (α1→6) linkage between the +1 and +2 subsites favors the formation of an (α1→3) linkage in the next reaction cycle; in turn, an (α1→3) linkage at this position induces the formation of an (α1→6) linkage. It has been proposed that the non-reducing end of the acceptor substrate is not well stabilized at the +1 site, while the glucosyl residue at the +2 site of ASR is more stabilized due to the stacking interaction with W675 and/or Y768 (88). Residue Y768 is unique for ASR and has been proposed to provide a 2nd stacking platform for a glucosyl residue at the +2 subsite. Thus, the linkage between the +1 and +2 glucosyl residues of a bound acceptor substrate probably determines that the +2 glucosyl residues stacks with either residue W675 or Y768, resulting in the accessibility of the C3 or C6 hydroxyl group of the non-reducing end glucosyl residue to the glucosyl-enzyme intermediate. Mutant Y768S:D769E:A770V of the alternansucrase of *L. mesenteroides* NRRL B-1355 was unable to synthesize alternan and produced more oligo-dextrans from sucrose and maltose (88). Another example is GTFA from *L. reuteri* 121 (73); characterization of the oligosaccharides initially produced by GTFA revealed that the most prominent products were oligosaccharides with alternating (α1→6) and (α1→4) linkages (111). However, GTFO from *L. reuteri* ATCC 55730, which shares high similarity with GTFA, synthesizes a reuteran with a high amount of (α1→4) linkages instead of alternating (α1→6) and (α1→4) linkages. Sequence alignment and structural analysis revealed three regions that might be responsible for the difference of linkage specificity between GTFA and GTFO. These three regions include a loop (residues K970 to Q977, GTFA numbering) from domain B, residue H1063:A1064 and a loop (residues D1083 to S1088, GTFA numbering) from domain A. The loop from domain B contributes to shaping the groove above the +1 site while the loop from domain A locates close to the +2 site. Residues H1063: A1064 is located downstream of the catalytic acid/base E1059 in homology region III. Mutagenesis studies showed that the loop from domain B and the loop from domain A are important for alternating (α1→6) and (α1→4) linkages synthesis (X. Meng et al., unpublished data). Mutations in the loop of GTFO to the corresponding loop in GTFA shifted GTFO from mainly (α1→4) linkage synthesis to alternating (α1→6) and (α1→4) linkages synthesis. GTF180

from *L. reuteri* 180 synthesizes an α-glucan with (α1→6) and (α1→3) linkages (71). Structural analysis of this α-glucan showed that the α-glucan is built up with different lengths of isomalto-oligosaccharides, interconnected by single (α1→3) linkages (108). All (-)α-D-Glc*p*-(1→3)- units were found to be 6-substituted and no consecutive (α1→3) linkages were found (108). This result suggests that GTF180 only forms (α1→6) linkages after (α1→3) linkages. Taken together, these examples show that the linkage specificity is also determined by the linkage between the +1 and +2 acceptor substrate binding site. However, unlike the situation in ASR, no 2nd aromatic residue was found at subsite +2 in GTFA and GTF180; multiple mechanisms of alternating linkage formation thus occur.

Based on the information obtained by structural and mutation studies, the ultimate goal is the design of glucansucrase variants capable of synthesizing α-glucan with desired properties for particular applications. However, this requires further mutation, structural and molecular dynamics (MD) studies to fully understand the mechanism of glucansucrases. Nevertheless, enzyme engineering has been shown to be effective in production of various α-glucans. The diversity of α-glucan produced could probably be further expanded by combining glucansucrase with other enzymes like glycogen branching enzymes. Hyper-branched α-glucans have been successfully produced by using a biomimetic system containing amylosucrase from *Neisseria polysaccharea* and the branching enzyme from *Rhodothermus obamensis* (168). Amylosucrase only catalyzes the synthesis of (α1→4)-linked glucan polymers (121), while glucansucrases synthesize a much more variety of glucans. Combining the diversity of glucansucrase and the other glucan-synthesizing enzymes is prospective for producing novel glucans.

Taken together, the production of tailor-made α-glucans by enzyme engineering, reaction engineering and combining with the other enzymes is very promising, and holds great potential for industrial applications in the food, medicine and cosmetic sectors.

Scope of the thesis

Glucansucrases, belonging to the GH70 family, are capable of producing α-glucans with different structures, especially regarding glycosidic linkage

specificity. Sequence alignments involving family GH70 enzymes comparisons with related family GH13 enzymes, and mutational studies, have identified several regions that are important for linkage specificity, most of these regions being close to the glucansucrase catalytic site. The recently determined GH70 glucansucrase crystal structures, such as GTF180-ΔN in complex with the substrate sucrose and with the acceptor substrate maltose, revealed the importance of previously unidentified residues, shaping the acceptor binding site, both from the catalytic domain A but, interestingly, also residues from domain B. In addition, the crystal structures revealed an unexpected domain organization with two novel domains (IV and V). The 3D structural information facilitates a more effective semi-rational engineering of glucansucrase enzymes. This PhD thesis structurally characterized the initial products of glucansucrase from sucrose, to analyze the α-glucan biosynthetic process, investigated the functions of domain V in glucansucrase enzymes following its truncation, and studied the effects of mutations in residues close to the acceptor substrate binding subsites on linkage and reaction specificity by semi-rational engineering.

Chapter 1 reviews current knowledge of glucansucrase enymes of family GH 70. The biological distribution, primary structure, domain organization, 3D structure, catalytic mechanism, reaction and linkage specificity of glucansucrase enzymes are reviewed and discussed.

Chapter 2 presents the structural characterization of oligosaccharides initially formed from sucrose or malto-oligosaccharides (MOS, DP2-6) by GTFA of *L. reuteri* 121 in order to gain insights into reuteran synthesis. Incubations with sucrose only, acting as both donor and acceptor substrate, yielded linear gluco-oligosaccharides, which are elongations of sucrose with glucose units through alternating (α1→4) and (α1→6) linkages. MOS alone were poor substrates for GTFA. In the presence of both sucrose and MOS, GTFA elongated MOS (acceptor substrate) with glucose units from sucrose (donor substrate) with predominantly alternating (α1→4) and (α1→6) linkages.

Chapter 3 describes the effects of sucrose concentrations on the product distribution of GTFA of *L. reuteri* 121. The ratio of oligosaccharide and polysaccharide synthesis was directly proportional to the concentration of sucrose, indicating that the product size distribution is kinetically controlled. The linkage

specificity remained unchanged at different concentrations of sucrose. The oligosaccharides produced at high sucrose concentration (1.0 M) were isolated and structurally characterized, revealing the synthetic process from sucrose to polysaccharides.

Chapter 4 investigates the functions of domain V in glucansucrase GTF180 of *Lactobacillus reuteri* 180. The crystal structure of the N-terminally truncated GTF180 (GTF180-ΔN) revealed that the polypeptide chain follows a U shape course to form five domains, including domains A, B and C, which resemble those of family GH13 enzymes, and two additional and novel domains (domains IV and V), which are attached to the catalytic core. Truncation of domain V of GTF180-ΔN heavily impaired its polysaccharide synthesis resulting in increased production of longer oligosaccharides. Mutations L940E and L940F in GTF180-ΔNΔV, in residues that are involved in acceptor-substrate binding, partially restored its polysaccharide synthesis. These results demonstrate that the interactions of growing glucan chains with both domain V and acceptor substrate binding sites are occurring during polysaccharide synthesis. Domain V is not involved in determining the linkage specificity and the size of polysaccharide produced; GTF180-ΔNΔV produced a polysaccharide identical in size and structure with that of GTF180-ΔN.

Chapter 5 reports that residue L940 in domain B of GTF180-ΔN has a crucial role in the linkage and reaction specificity of this enzyme. Mutations in L940 of GTF180-ΔN all caused an increased percentage of (α1→6) linkages and a decreased percentage of (α1→3) linkages in the products. α-Glucans containing 67% to 100% of (α1→6) linkages were produced by GTF180 and its L940 mutants. Particularly, mutant L940W was unable to form (α1→3) linkages and synthesized a smaller and linear glucan polysaccharide with only (α1→6) linkages.

Chapter 6 characterizes the effects of mutations in several residues located close to the +1 acceptor binding subsite (D1028 and N1029 from domain A, as well as L938, A978 and L981 from domain B) on reaction and linkage specificity of GTF180-ΔN. Our data showed that all these residues, except for L981, are critical for linkage specificity. Mutating A978 and D1028 to amino acid residues with

larger side chains resulted in a decrease in branched linkage formation probably due to steric hindrance effects caused by these bulky residues in a groove above the +1 acceptor binding subsite. D1028 mutants showed increased (α1→6) linkages synthesis as well. In view of the high hydrolysis/transglycosylation ratio of N1029 mutants, this residue is essential for transglycosylation activity, suggesting that the the hydrogen bond of N1029 with the carbohydrate acceptor substrate play a critical role.

Chapter 7 reports the production of novel hyper-branched α-glucans from sucrose by engineered glucansucrase GTF180 enzymes. In this chapter, residues D1085, R1088 and N1089 in GTF180 of *L. reuteri* 180, which all form an indirect hydrogen bond with the +2 glucosyl unit of maltose through the same water molecule, were targeted for mutagenesis. All combined mutants, and the D1085 and R1088 single mutants, showed increased percentages of branching in the α-glucans produced (from 13% in wild-type GTF180-ΔN upto 22%). These mutant α-glucans also possessed a small amounts of (α1→4) linkages (at the most 5%). Single residue analysis showed that mutations in D1085 and R1088 (but not N1089) are responsible for the increase in branched linkage formation.

Chapter 8 presents the summary of the thesis and concluding remarks.

Table S1. Detailed information of glucansucrase enzymes used in the phylogenetic analysis

Enzyme	Strains	GeneBank	Polysaccharide	Reference
GTF1971	*Lactobacillus animalis* TMW 1.971	CCK33644.1	dextran	(169)
GTF1624	*Lactobacillus curvatus* TMW 1.624	CCK33643.1	dextran	(169)
GTFKG3	*Lactobacillus fermentum* KG3	AAU08008.1	dextran	(71)
GTF33	*Lactobacillus parabuchneri* 33	AAU08006.1	dextran	(71)
GTF180	*Lactobacillus reuteri* 180	AAU08001.1	dextran	(71)
GTFML1	*Lactobacillus reuteri* ML1	AAU08004.1	mutan	(71)
GTFA	*Lactobacillus reuteri* 121	AAU08015.1	reuteran	(70)
GTFO	*Lactobacillus reuteri* ATCC 55730	AAY86923.1	reuteran	(74)
GTF106A	*Lactobacillus reuteri* TMW 1.106	ABP88726.1	dextran with 15% (α1→4)	(170)
GTFB	*Lactobacillus reuteri* 121	AAU08014.2	4, 6-α-glucanotransferase	(79)
GTFML4	*Lactobacillus reuteri* ML1	AAU08003.2	4, 6-α-glucanotransferase	(80)
GTFW	*Lactobacillus reuteri* DSM 20016	ABQ83597.1	4, 6-α-glucanotransferase	(80)
GTFKg15	*Lactobacillus sakei* KG15	AAU08011.1	dextran	(71)
LcDS	*Leuconostoc citreum* HJ-P4	BAF96719.1	dextran	(171)
DSRF	*Leuconostoc citreum* B/110-1-2	ACY92456.2	dextran	(54)
Alternansucrase	*Leuconostoc citreum* ABK-1	AIM52834.1	ND	---
DexT	*Leuconostoc citreum* KM20	ACA83218.1	ND	(172)
Glucansucrase EG001	*Leuconostoc lactis* EG001	ACT20911.1	ND	(173)
DSRA	*Leuconostoc mesenteroides* NRRL B-1299	AAB40875.1	dextran	(50)
DSRB	*Leuconostoc mesenteroides* NRRL B-1299	AAB95453.1	dextran	(51)
DSRS	*Leuconostoc mesenteroides* NRRL B-512F	AAD10952.1	dextran	(85)
DSRT	*Leuconostoc mesenteroides* NRRL B-512F	BAA90527.1	dextran	(91)
DSRC	*Leuconostoc mesenteroides* NRRL B-1355	CAB76565.1	dextran	(52)
ASR	*Leuconostoc mesenteroides* NRRL B-1355	CAB65910.2	alternan	(52)
DSRB742	*Leuconostoc mesenteroides* B-742B	AAG38021.1	dextran	(174)
DSRD	*Leuconostoc mesenteroides* LCC4	AAG61158.1	dextran	(175)
DSRE CD1	*Leuconostoc mesenteroides* NRRL B-1299	CAD22883.1	dextran	(47)
DSRE CD2	*Leuconostoc mesenteroides* NRRL B-1299	CAD22883.1	(α1→2)	(47)
DSR-DP	*Leuconostoc mesenteroides* NRRL B-1299	CDX66641.1	dextran	(36)
BRS-A	*Leuconostoc mesenteroides* NRRL B-1299	CDX66896.1	(α1→2)	(36)
DSR-M	*Leuconostoc mesenteroides* NRRL B-1299	CDX66895.1	dextran	(36)
DSRR	*Leuconostoc mesenteroides* NRRL B-1501	AAN38835.1	ND	---
DSRP	*Leuconostoc mesenteroides* IBT-PQ	AAS79426.1	dextran	(176)
DSRX	*Leuconostoc mesenteroides* L0309	AAQ98615.2	ND	---
DexYG	*Leuconostoc mesenteroides* 0326	ABC75033.1	dextran	(177)
DSRBCB4	*Leuconostoc mesenteroides* NRRL B-1299CB4	ABF85832.1	dextran	(178)
DSRN	*Leuconostoc mesenteroides* KIBGE IB-22	AFP53921.1	dextran	(179)
GTFI	*Streptococcus criceti* GTC242/HS-6	BAF62338.1	mutan	(180)
GTFS	*Streptococcus downei* MFE 28	AAA26898.1	dextran	(67)
GTFI	*Streptococcus downei* MFE 28	AAC63063.1	mutan	(87)
GTFG	*Streptococcus gordonii* str.Challis substr.CH1	AAC43483.1	dextran	(181)
GTFD	*Streptococcus mutans* UA159	AAN58619.1	dextran	(182)
GTFB	*Streptococcus mutans* UA159	AAN58705.1	mutan	(183)
GTFC	*Streptococcus mutans* UA159	AAN58706.1	mutan	(183)
GTFR	*Streptococcus oralis* ATCC10557	BAA95201.1	dextran	(184)
GTF14035	*Streptococcus orisuis* NUM 1001/JCM14035	BAF62337.1	mutan	(180)
GTFJ	*Streptococcus salivarius* ATCC 25975	AAA26896.1	mutan	(185)
GTFK	*Streptococcus salivarius* ATCC 25975	CAA77898.1	dextran	(185)
GTFL	*Streptococcus salivarius* ATCC 25975	AAC41412.1	alternan	(185)
GTFM	*Streptococcus salivarius* ATCC 25975	AAC41413.1	dextran	(185)
GTFP	*Streptococcus sanguinis* ATCC 10556	BAF43788.1	dextran	(186)
GTF-I	*Streptococcus sobrinus*	BAA14241.1	mutan	(187)
GTFI	*Streptococcus sobrinus* ATCC 33478/OMZ176	BAA02976.1	mutan	(188)
GTFU	*Streptococcus sobrinus* B13N	BAC07265.1	Highly branched	(189)
GTFT	*Streptococcus sobrinus* B13N/OMZ176	AAX76986.1	ND	---
DSRWC	*Weissella cibaria* CMU	ACK38203.1	dextran	(84)
DSRK39	*Weissella cibaria* LBAE-K39	ADB43097.3	dextran	(82)
WcCab3-DSR	*Weissella confusa* Cab3	AKE50934.1	dextran	(190)
DSRC39-2	*Weissella confusa* LBAE C39-2	CCF30682.1	dextran	(191)
WcE392-DSR	*Weissella confusa* VTT E-90392	AHU88292.1	dextran	(191)

ND: not determined